DOCUMENT RESUME

ED 105 479 CS 202 007

AUTHOR Reynolds, Allan G.
TITLE Coding, Ordering, and Retrieving Information from Sentences. Final Report.
INSTITUTION Dartmouth Coll., Hanover, N.H.
SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
BUREAU NO BR-I-0174-A
PUB DATE Nov 72
GRANT OEG-1-71-0101(508)
NOTE 84p.

EDRS PRICE MF-$0.76 HC-$4.43 PLUS POSTAGE
DESCRIPTORS *Cognitive Processes; *Educational Research; English; Grammar; Higher Education; *Information Retrieval; Language Skills; Memory; *Recall (Psychological); *Sentence Structure

ABSTRACT

Four experiments are reported which examine the role of phrase structure, memory load, concreteness of materials and other variables in the recall of meaningful English sentences. Several major findings are reported. Concreteness of the stimulus materials consistently is an aid to recall; although this is predicted from an imagery interpretation of the coding of concrete materials, some of the results are at odds with such a theory. Children as young as six years seem to chunk sentences for recall essentially the same as adults; however, the storage systems seemed to differ. No relation to functional knowledge of grammar was found. Results from two experiments question the adequacy of a theory that depends heavily on surface structure of the sentence as a medium for storage. The general discussion proposes that at least four interactive processes are necessary to explain the obtained recall results. (Author)

ED105479

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Final Report

Project No. 1-0174-A
Grant No. OEG-1-71-0101(508)

Allan G. Reynolds
Dartmouth College
Hanover, New Hampshire 03755

CODING, ORDERING, AND RETRIEVING INFORMATION FROM SENTENCES

November, 1972

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development

CS 202 007

Abstract

Four experiments are reported which examine the role of phrase structure, memory load, concreteness of materials and other variables in the recall of meaningful English sentences. Several major findings are reported; concreteness of the stimulus materials consistently is an aid to recall; although this is predicted from an imagery interpretation of the coding of concrete materials, some of the results are at odds with such a theory. Children as young as six years seem to chunk sentences for recall essentially the same as adults; however, the storage systems seemed to differ. No relation to functional knowledge of grammar was found.

Results from two experiments question the adequacy of a theory that depends heavily on surface structure of the sentence as a medium for storage. The general discussion proposes that at least four interactive processes are necessary to explain the obtained recall results.

Final Report

Project No. 1-0174-A
Grant No. 1-71-0101(508)

Coding, ordering, and retrieving information from sentences

Allan G. Reynolds
Dartmouth College
Hanover, New Hampshire 03755

November, 1972

The research reported herein was performed pursuant to a grant from the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express fully their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development

I wish to acknowledge the kind assistance of the Hartford (Vt.) School Board, and Mrs. Constance Reid, and the teachers and pupils of the Wilder (Vt.) Elementary School.

Table of Contents

List of Tables i

List of Figures ii

Introduction 1

Experiment 1 8
Method 8
Results and Discussion 13

Experiment 2 32
Method 33
Results and Discussion 34

Experiment 3 40
Method 40
Results and Discussion 41

Experiment 4 49
Method 50
Results and Discussion 53

General Conclusions 67

Footnotes 70

References 71

List of Tables

		Page
Table 1-1.	Sentences used in study one.	11-12
Table 1-2.	Number of sentences entirely correct, interaction of Phrase Position by Concreteness.	15
Table 1-3.	Mean proportion of words recalled correctly, interaction of Phrase Position by Concreteness.	17
Table 1-4.	Mean forward deletion TEPs, interaction of Phrase Position by Concreteness.	19
Table 1-5.	Mean proportion of words recalled as omission errors, interaction of Phrase Position by Concreteness.	23
Table 1-6.	Mean forward stop TEPs, interaction of Concreteness by Transition Type.	27
Table 1-7.	Mean forward deletion TEPs, interaction of Transition Type by Number of Adjectives.	28
Table 2-1.	Mean proportion of words recalled correctly, interaction of Grade by Concreteness.	35
Table 3-1.	Proportion of sentences recalled entirely correct, interaction of Digit Length by Clause Position by Concreteness.	44
Table 4-1.	Proportion of times recognized correctly at various confidence levels.	54
Table 4-2.	Mean proportion correct (NEW phrases), interaction of Cue Position by Type of Change.	56
Table 4-3.	Mean proportion correct (NEW phrases), interaction of Part of Speech Changed by Cue Phrase Presented.	58
Table 4-4.	Mean proportion correct (NEW phrases), interaction of Part of Speech Changed by Type of Change.	60
Table 4-5.	Proportion of words recalled correctly, interaction of Concreteness by Voice.	63

List of Figures

		Page
Figure 1-1.	Number of sentences entirely correct, interaction of Phrase Position by Concreteness.	16
Figure 1-2.	Mean proportion of words recalled correctly, interaction of Phrase Position by Concreteness.	18
Figure 1-3.	Mean forward deletion TEPs, interaction of Phrase Position by Concreteness.	20
Figure 1-4.	Mean proportion of words recalled as omission errors, interaction of Phrase Position by Concreteness.	24
Figure 1-5.	Mean forward deletion TEPs, interaction of Transition Type by Number of Adjectives.	29
Figure 2-1.	Mean proportion of words recalled correctly, interaction of Grade by Concreteness.	36
Figure 3-1.	Proportion of sentences recalled entirely correct, interaction of Digit Length by Clause Position by Concreteness.	45
Figure 4-1.	Mean proportion correct (NEW phrases), interaction of Cue Position by Type of Change.	57
Figure 4-2.	Mean proportion correct (NEW phrases), interaction of Part of Speech Changed by Cue Phrase Presented.	59
Figure 4-3.	Mean proportion correct (NEW phrases), interaction of Part of Speech Changed by Type of Change.	61
Figure 4-4.	Proportion of words recalled correctly, interaction of Concreteness by Voice.	64

Introduction

One of the most remarkable human abilities is the capacity to remember things; that is, to store complicated information. There is an entire body of research literature in verbal learning examining the parameters and processes of free recall of words, paired-associate learning, and serial learning of lists of items. However, people more typically use their language abilities in the form of sentences; paradoxically, learning and memory research involving sentences is less widespread than that involving discrete words.

Sentences are typically remembered better than strings of words (e.g., Marks & Miller, 1964); hence, the question arises: "What do people remember when they remember sentences?" It is an everyday observation that people often remember the essence of a sentence rather than the specific words of the sentence; as well, there is much experimental evidence to substantiate this view. Binet and Henri (1894) found that the specific characteristics of a sentence were not retained in memory. Mehler's (1963) Ss often confused active and passive voice (in recall), but most of their mistakes were "meaning-preserving" errors. Mehler suggested that Ss remember the kernel string and the transformational markers of a sentence independently and that they are more likely to forget the markers. However, Fillenbaum's (1966) Ss made lexical not transformational mistakes while preserving the underlying meaning. As well, Sachs (1967a, 1967b) and Wanner (1968) report that Ss are better able to recognize changes in meaning than changes in wording alone. It must be noted that transformational changes also entail semantic changes. Thus, do Ss remember transformational changes or the covarying semantic changes? Even though actives and passives are logically synonymous, they are used for different purposes (Clark, 1965; Halliday, 1967a, 1967b; Hebb, 1969; Johnson-Laird, 1968; Morton, 1966; Prentice, 1967; Rommetveit, 1968; Tannenbaum & Williams, 1968; Turner & Rommetveit, 1968). Thus, it would seem that semantic aspects are more important than syntactic ones.

Why then is grammar important or functional in the learning of sentences? It is probably because grammar gives an organization to language that is not present in a string of words. As Mandler (1967) has suggested, not only are memory and organization correlated, but organization is a prerequisite to memory. It has been posited that people use "meaningful units" in sentences as memorial aids (cf. Thorndike's (1931) concept of "belongingness"). Johnson (1968a) has proposed that "grammar provides Ss with a ready-made and reliable recoding scheme" (p. 437). The meaningful units in sentences seem to functionally resemble the "chunks" of information proposed by Miller (1956a, 1956b).

Miller proposed a limit of 7 (±2)on human short-term memory capacity. Since we can remember more than seven pieces of information, Miller used the Unitization Hypothesis for an explanation and proposed chunking as a method to increase storage capacity. We could not simply add seven more items of information to our memory for we would then lose the first seven. For storage capacity to increase, each remembered item must be enriched with more information. Thus, chunks are "informationally-rich items"; Miller thus used the idea of chunking in two senses: one was the grouping of information into familiar units (chunks). The second use involved the process of chunking chunks to form larger and larger chunks of information.

In sentences, what defines the nature of a chunk? Much experimental evidence seems to point toward phrases or phraselike units as behaving much like chunks. Maclay and Osgood (1959) found that pauses in speech occur before lexically difficult items and at syntactic junctions, and that repetitions characteristically involve small phrase units. Miller (1963, in Garrett & Fodor, 1968) showed that when Ss copy text, they "hold" a phrase in memory, then consult the text again. Also, a series of experiments from MIT (Fodor & Bever, 1965; Garrett, Bever, & Fodor, 1966) showed that Ss displace "clicks" imbedded in sentences toward major phrase boundaries. Anglin and Miller (1968) found that paragraphs presented in phrase segments were remembered better than those presented with no phrase segmentation. Marks (1967) has shown that sentence inversions that interrupted major phrase boundaries had longer source-identification times than those with phrase boundaries intact. This is similar to Anglin and Miller's (1968) finding that paragraphs segmented at phrase boundaries had higher recall scores than paragraphs where the segmentation did not correspond to the phrase boundaries. Epstein (1961, 1962, 1963) found that word endings indicative of phrase organization aided memorization; Lachman and Tuttle (1965) further specified the locus of this effect as being in the storage of the sentence not in perception or recall. Epstein (1967) segmented syntactically structured but anomalous material at phrase junctures; the cues that helped to chunk material also aided retention. Johnson (1965) showed that most errors of recall occurred at phrase boundaries indicating that Ss were using their knowledge of grammar to segment a sentence into functional subunits.

There appears to be much evidence that syntax affects sentence retention by defining the phrase structure of the sentence. This in turn determines into how many chunks the sentence is broken. The retention of phrases might correspond to the "adopted chunks" noted by Tulving and Patkau (1962); they found that larger chunks could be

constructed with higher approximations to English (cf. also McNulty, 1966). Epstein (1967), on the other hand, has suggested that chunking is facilitating for learning sequences only when some high degree of syntactic structure is present.

The basic position being offered here is that in memory for sentences, what is stored are chunks whose boundaries are derived from the surface structure of a sentence. As Wearing (1970) notes, chunks may be nested within one another (cf. Miller's "super-chunks"). Wright (1968) has raised the possibility that we can conceptualize sentences in terms of both the number of chunks into which they are broken and the number of items of information (words) in each chunk. Thus, a sentence such as "The old man thanked the boy" could be described as a 3-1-2 sentence. But we cannot assume uniform difficulty of the items in a chunk. Roberts (1967) stated that nouns in a sentence were less subject to forgetting than adjectives or adverbs. Matthews (1968), using Savin and Perchonock's (1965) technique, found that Ss seem to encode sentences by a complete encoding of syntactic forms with the emphasis on noun and verb components and subsidiary and less efficient storage of qualifiers. This suggests that nouns and verbs are chunked or recoded first, and various qualifiers are chunked separately or even discarded if memory load is too high. Martin, Roberts, and Collins (1968) claim that when a sentence is learned, its grammatical structure is determined, and then words are differentially processed into memory. The critical variable is their importance for the sentence. Perfetti (1969) offered a somewhat different approach to this issue. In order to understand his argument, it is first necessary to digress a little. Martin and Roberts (1966), using Yngve's (1960) theory, proposed that the average depth ($\bar{d}$) of a sentence was an important factor in sentence retention. This was substantiated by Martin, Roberts, and Collins (1968) using single sentence recall. However, Martin and Roberts (1967) found no effect for $\bar{d}$ in recall. This led Perfetti to propose that it was not average depth but number of lexical words that was important. In fact, he proposed it was the ratio of the number of lexical words to the total number of words that was of special importance. By his definition, lexical words included nouns, verbs, adjectives, etc. This study also showed nouns to be of greater importance in retention. In an unpublished study, Mlle. Maxianne Berger and I have found converging evidence using bilingual (French-English) Ss. These Ss had to translate sentences mixed in two languages (Franglais) (e.g., "The quiet priest thanked the homme French"). The results indicated that primary attention was paid to the language and position of the nouns.

A recent paper (Begg & Paivio, 1969) has introduced a

new factor in this argument. They propose that there are actually two coding mechanisms operating. Which system is used depends largely on the nature of the stimulus material. If the sentences are concrete, nonverbal images are evoked (cf. Rommetveit & Turner, 1967); thus, concrete material can be stored in parallel by nonverbal spatial imagery rather than sequentially, as in strings of words. Information contained in abstract material, on the other hand, is linked more closely to the sequentially organized verbal units themselves. Therefore, with concrete material the meaning of the whole sentence is stored, but with abstract material specific words and phrases are retained. Remember, Sachs (1967a, 1967b) and Wanner (1968) have reported that Ss could recognize changes of meaning better than changes of words; both had used concrete material. Begg and Paivio predicted that in a recognition paradigm, Ss would be able to easily recognize meaning changes in concrete sentences but would find abstract material easier when spotting lexical changes. Their results agreed with this prediction.

There are important consequences of such a dual-coding theory. If concrete sentences are coded as images, the order of ease of recall should be nouns (easiest), adjectives, verbs, then adverbs. There is more semantic distortion as we progress from nouns to adverbs; adverbs have less power in establishing concrete sentential meaning. We may predict from this that nouns should function as the best retrieval cues (Horowitz & Prytulak, 1969; Lambert & Paivio, 1956; Paivio, 1963). Another prediction of this theory would be that voice should be often confused in recall of concrete sentences but not in abstract. Support for the dual-coding hypothesis has come from a number of sources. Sasson and Fraisse (1972), using a retroactive-interference paradigm, found that interpolated concrete sentences and pictures produced equal amounts of significant interference for recall of concrete sentences. Recall of abstract sentences was not affected by interpolated concrete sentences or pictures but was affected by interpolated abstract sentences. Sasson and Fraisse view their results as supporting the hypothesis that the contents of concrete sentences are stored primarily as visual images. Philipchalk (1972), exploring Pompi and Lachman's (1967) notion of a surrogate structure as a substrate of connected discourse, asked Ss to recall concrete or abstract paragraphs that differed on degree of thematicity (thematic or random order). He found that Ss were most adept at recall when materials were thematic and concrete. This finding obtained also after a two-week delay. Philipchalk interpreted this to mean that concrete thematic discourse has a unique property not shared by either abstract material or by non-thematic concrete discourse. Of course, Philipchalk's manipulation of the thematicity of the

paragraphs also disrupted an important set of cues for organized retention that have already been discussed--the phrase-structure divisions of the sentences. Nonetheless, it appears that disruption of these cues is more harmful for concrete than for abstract sentences. Following Epstein (1967), who suggested that chunking is a valuable aid only when some degree of syntactic structure is present, it may be that concrete sentences (when syntactic structures are present) are most aided by use of chunking strategies. If memory for abstract sentences is more tied to the left-to-right order of words and relies more on word-to-word memory than global memory, randomization of the order of words may not be removing much of the information useful for recall. With concrete sentences on the other hand, the chunking may be of a quite different nature. If a surrogate structure is formed for thematic concrete discourse (image or schemata), then removal of the phrase-structure chunking cues which permit the discovery of the large units which are the components of the surrogate structure will obviously impair the formation of the image.

It must be kept in mind that there may be many possible ways for chunking to occur. Chunking may be induced or facilitated by the nature of the materials to be learned, by various experimenter manipulations, and on Ss's own initiative. Epstein (1967) has suggested that a certain level of syntactic organization is essential to the use of chunking with verbal sequences. This leads to our present contention that the phrase-structure of a sentence offers a ready-made set of chunks and chunk boundaries for Ss. Of course, Johnson (1970), by using physical spaces between letter groupings, has also presented ready-made chunks to his Ss.

As well, a number of other manipulations have been used by experimenters to induce chunking. McLean and Gregg (1967) also used letter groupings to induce chunking. Loughery and Spector (1972) varied number of repetitions of groups of letters and found this to be a significant factor in recall.

Finally, it is self-evident from everyday observation that we all attempt to chunk information when memory load would be heavy. Seven-digit phone numbers are usually rehearsed and communicated in chunks of 3 and 4 digits. Thompson and Roenker (1971) have even presented an analysis of how Ss "learn" a clustering strategy in an experimental task.

In summary, the research to date would suggest that the facilitating effect of grammatical organization on sentence learning lies mainly in its organizing and segmenting properties. There are still many questions unanswered concerning the processes by which coding of sentential information takes place. If we assume that "chunking" of

some sort does take place in storage of sentences, several major questions may be asked:

1. How are chunks labelled or identified in memory?
2. How are chunks retrieved from memory storage?
3. How are chunks kept ordered in memory?
4. How does recall of one chunk affect the probability of recall of others?
5. What is the developmental sequence that gives rise to the ability to chunk verbal information?

The most general, and ambitious, attempt to answer these kinds of questions is provided by the decoding-operation model of N. F. Johnson (1970, 1972). This model assumes that there is a single memorial representation (code) for the information contained in a chunk. As a corollary of this view, the model also assumes that there are no direct connections among the items in a chunk of information. Each item is independently tied to the chunk's code.

This view of coding makes several firm predictions about learning and recall. If chunks are unitary, they should be recovered from memory in an all-or-none fashion. Once a chunk code is recovered from memory, all the information represented by a code should be available. Since codes may represent either other codes or response items, there exists the possibility of hierarchical ordering of sequential information.

Johnson's decoding-operation model hypothesizes that (at the time of recall) S recovers some code for an entire sequence and then decodes it into the codes (or items) at the next lower level of organization in the hierarchy. It is assumed that the codes recovered in each decoding operation are tagged with order information. A S will temporarily store all codes except for the one with temporal priority (information gathered from the order tag); this code will be decoded and so on until a response item is produced. At this time S will retrieve another code from memory and continue the process. An interesting possibility raised by the model is that the decoding and "read-out" functions may be performed more or less in parallel; S may be producing item X while decoding the code that represents item X + 1. A final assumption of the model is that Ss are relatively conservative and will terminate their recall attempts when uncertainty about a code arises.

If Ss make recall decisions at chunk boundaries (codes are decoded here), it is at these points that Ss discover their uncertainty. Experimental results (e.g., Johnson, 1970) indicate that there are clear TEP spikes on the transitions between the last item of one chunk and the first item of the next chunk.

As well, Johnson (1970) has found that the probability of terminating a recall attempt (as measured by the stop

TEP) increases with increase in the size of the following chunk. Johnson (1970) has also shown that probability of terminating a recall attempt is a function not of how many items are left to be recalled but rather a function of how many chunks of information remain.

A further set of predictions derived from Johnson's model concerns the relationships between items and between an item and the chunk code. Chunks are assumed to function like "opaque containers." If any item is changed, the chunk code would no longer apply. Results from several studies (Johnson, 1970)confirm that changing a letter in a chunk results in large losses of information regarding both the changed and unchanged letters. As well, the degree of mutilation of a chunk seems to have no effect: changing one letter is as effective in disrupting recall as changing two or three letters.

The last problem to be tackled by the model is that of storage of order information. The hierarchical organization of information imposes some restraint on the number of ordered sequences that could be recalled. Assume a seven-item sequence (AB/CD/EFG). Let the slashes signify the boundaries between chunks. If S is in the process of decoding the AB chunk, it is impossible for him to recall G since G is not available for recall at this time. This severely limits the number of orders which can be recalled but an additional mechanism is necessary. Johnson (1972) assumes the existence of "order tags" associated with each item or code. The presence of order tags for A and B permits S to recall the items in this chunk in the correct order, that is, A precedes B. Johnson hypothesizes that each item may be tagged for order all the way up the organizational hierarchy; he presents data (Johnson, 1970) to support this view.

Johnson's model is the most explicit current attempt to empirically test questions and predictions arising from a coding view of human memory. In the four experiments to follow some aspects of the model will be tested in various experimental situations. Unlike the majority of Johnson's studies (which used letter sequences) our experiments will use meaningful English sentences where the chunk divisions will be assumed to be defined by the phrase structure of the sentences.

Experiment 1

As a logical outgrowth of Miller's (1956a, 1956b) chunking proposal, it may be asked if recall of chunks is "all-or-none" (called the "initial reproducing tendency" by Muller and Pilzecker, 1900). Johnson (1969) has suggested that chunks function like "opaque containers." His study supports the conception of each element of a chunk being separately tied to the chunk code itself; the alternative view of strong intra-chunk associations among elements (with one comprehensive chunk code) was not supported. However, Johnson used letter trigrams rather than sentence elements. This ignores both preexisting associations among words and the constraints imposed by grammatical structure. If grammar functions in the ways we have intimated, the relations among words in a sentence may affect the "all-or-none" process. As Johnson (1968b) himself has found, phrases and phrase-like strings tend to be recalled more often in an all-or-none fashion than random strings of words.

Four main variables are expected to affect all-or-none recall of phrases:

1. Sentence concreteness--Following Begg and Paivio's (1969) reasoning, we might expect more all-or-none recall for the sentence as a whole for concrete sentences.
2. Adjective deletion--If size of a chunk is a factor in all-or-none recall, deletion of one or more adjectives should facilitate all-or-none recall.
3. Adverb phrase position--The closer the adverb phrase is to the verb, the more likely its recall with the verb should be.
4. Association value--Elements linked with high association value should be more likely to be recalled together than those with low value.

There are several other clear predictions from Johnson's model. If Ss are in fact treating phrases as chunk-like units in the present study, there should be clear TEP spikes at phrase boundaries. If, as was found in Johnson (1969), the stop TEP at the transition between the first and second chunks is only affected by the number of remaining chunks, there should be a clear effect for adverb phrase position. Finally, if the size of the first chunk is critical in determining number of omission errors, then the size of the first chunk should affect number of omission errors for the sentence.

Method

Subjects. A total of 179 students from the Introductory Psychology course at Dartmouth College served as subjects; thirty-six participated in the experimental sessions, and

the remaining 143 Ss helped in preparation of the response sentences.

Design. A paired-associate paradigm was employed. The stimuli were two-digit numbers[1]; the responses were sentences. The experiment was run as a factorial design with two independent factors and two repeated-measures factors. The independent factors were Concreteness (sentences were either concrete or abstract) and Association Value of the adjectives in the subject phrase (high or low association with the subject noun); the repeated measures were Number of Adjectives presented in the subject phrase (two[2], one[3], or zero) and Phrase Position (the adverbial phrase could occur at the beginning or end of the sentence, or be absent from the sentence entirely). The design was thus a 2 x 2 x 3 x 3 factorial design with repeated measures on the last two factors. Nine Ss were tested in each of the four cells formed by the two independent factors.

Materials. The sentences had the basic form Article, Adjective 1, Adjective 2, Subject Noun, Verb, Article, Object Noun, Preposition, Article, Phrase Noun (e.g., The weary old man opened the umbrella near the park). Nine concrete sentences and nine abstract sentences were constructed using the following procedures. A pool of 200 abstract sentences, having the general form described above, except for the adjectives, was generated by the experimenters. Four judges evaluated every sentence on the criteria of adequate abstractness of each word and the whole sentence, and meaningfulness of the whole sentence when the adverb phrase was in each of the possible positions. Any sentence questioned by any one of the judges on these criteria was discarded; 65 sentences remained in the pool. These 65 sentences (which lacked adjectives in the subject phrase) were presented to 83 volunteer subjects from Introductory Psychology. These subjects were instructed to supply two adjectives that would "very commonly" appear with the subject noun in each sentence. These adjectives were then subjected to a frequency count, and the two adjectives most frequently used with each sentence were selected as the high association (HA) adjectives.

Since some attempt would be made to compare recall of various parts of a sentence (adjective, noun, verb, etc.), it was felt that those parts should be matched on Thorndike-Lorge frequency as closely as possible. To accomplish this, the 65 sentences (including the HA adjectives) were sorted by a computer program which endeavoured to find the nine sentences with the greatest similarity of Thorndike-Lorge frequency among the words in the sentences. The program selected nine sentences at random from the pool of 65

sentences, computed the average Thorndike-Lorge frequency (by rank in the 54 words in temporary store) for the adjectives, the nouns, the verbs, etc., and then computed the variance of these means. This process was reiterated 100 times, and the sample of nine sentences with the lowest variance of ranked frequencies was chosen as the stimulus group. These nine sentences were then examined to be sure that no adjective was used in more than one sentence (one adjective had to be replaced by the next most frequently given adjective). For each sentence two adjectives that had a frequency of one and matched the two HA adjectives on Thorndike-Lorge value were selected as low association (LA) adjectives.

Once the nine abstract sentences were generated, these were used as the models for construction of nine concrete sentences matched word-for-word with the abstract sentences on Thorndike-Lorge values. A computer program generated from pools of stored verbs, nouns, and adverbs a sample of sentences which matched each of the abstract sentences for Thorndike-Lorge value within a range of ±5. The adjectives were not included in this generation. Nine concrete sentences which were judged to be meaningful and realistic were chosen from the generated pool, and these were submitted to a sample of 60 Ss from Introductory Psychology who supplied "common" adjectives as was done for the abstract sentence pool above. These adjectives were counted for frequency of use in the nine sentences, and the two in each sentence that were the most commonly used and had the closest approximation to the abstract adjectives in Thorndike-Lorge value were selected as HA concrete adjectives. The two adjectives with the lowest usage which matched the corresponding abstract adjectives were designated LA concrete adjectives. The concrete and abstract sentences that were finally chosen are presented in Table 1-1.

Insert Table 1-1 about here

Procedure. Testing of Ss took place at a computer teletype connected to the Dartmouth Time-Sharing System; Ss were prevented from seeing what was being printed by a small screen taped to the teletype. When an entire line had been printed, this was advanced into view so that S could see it all at once. After Ss were seated at the teletype and had read the instructions, they were given a presentation trial consisting of nine number-sentence pairs with a 4-sec inter-pair interval. After a 15-sec pause, the first test trial began. The teletype printed one of the two-digit numbers, and S was to input the associated sentence. After S had input an answer (or "DON'T KNOW"), the computer asked S to

Table 1-1. Concrete and abstract sentences used in the study
(low association adjectives are in parentheses)

Concrete

The faithful, Indian (bold,mountain)scout led the cattle through the gap.

The concerned, loving (protective,helpful) parent unpacked the cheese after the almonds.

The rich, greedy (fat,aged) king polished the diamond for the men.

The small, sharp (single, thick)thorn jabbed the minister in the wrist.

The frightened, little (helpless,lost) child halted the colt by the bridge.

The tall, old (strong,young) sentry guarded the wharf near the rocket.

The alert, impartial (deliberate,exacting) referee identified the winner after the race.

The proud, diligent (beaming,friendly) farmer wiped the tractor before the market.

The daring, radical (indignant,alert) student protected the beggar from the porter.

Abstract

The careful, thorough (severe,second) review missed the error despite the effort.

The stiff, pompous (somber,drab) formality characterized the occasion at the outset.

The recent, bad (dark,tragic) experience obscured the issue before the exploration.

The bitter, suppressed (pure,wild) hate ended the harmony during the estrangement.

The revolutionary, political (strong,conservative) movement derived the power from the incident.

The repeated, futile (wasted,tiring)attempts caused the

frustration for a while.

The radical, new (rhetorical,political) ideology won the support after the reform.

The sound, clear (strict,concise) logic outdid the intuition in the session.

The stupid, careless (needless,incredible) blunders plagued the theory throughout the debate.

verify that this was indeed the answer he wished to enter. After S verified his input, the correct answer was presented. After a 4-sec pause, another two-digit number was presented, and S made another response. This testing procedure continued until all nine sentences had been tested for recall.

Ss were given six presentation and six test trials on the group of nine sentences. The order of presentation of sentences was randomized on each trial. Following these six trials, each S completed a questionnaire to assess mnemonic strategies.

Results and Discussion

Whole Sentence Analyses

A four-way analysis of variance design utilizing the factors of Concreteness (concrete and abstract sentences), Association Value of the adjectives (high or low), Number of Adjectives in the subject phrase (two, one, or none), and Phrase Position (end of sentence, beginning of sentence, or absent) was used for a number of initial analyses. The first two factors were based on independent groups while the second two factors were repeated measures. The dependent measures analyzed in this design included a) number of times the sentence was entirely correct, b) proportion of words in the sentence recalled correctly, c) average transition error probabilities (backward and forward TEPs) in the sentence of three kinds--deletion, commission, and stop (Johnson, 1970), d) proportion of errors of each of three kinds--synonym errors, omission errors, and commission errors (intrusion errors).

Number of times sentence entirely correct. Concrete sentences were recalled completely correctly more often on the average than abstract sentences (2.9 vs 1.63, $F = 12.66$, $df = 1/32$, $p < .01$). As well, a significant effect ($F = 14.55$, $df = 2/64$, $p < .01$) was obtained for Number of Adjectives; two-adjective sentences ($\bar{X} = 1.7$) were recalled completely correct less frequently than either one-adjective sentences ($\bar{X} = 2.45$) or zero-adjective sentences ($\bar{X} = 2.62$). There was no significant difference between one-adjective and zero-adjective sentences. The most parsimonious explanation of this result is that the two-adjective case simply presents more items for S to remember and hence poorer recall. A similar interpretation may be applied to a significant result ($F = 24.73$, $df = 2/64$, $p < .01$) for Phrase Position. When the adverb phrase is absent ($\bar{X} = 2.94$), recall of the entire sentence is better than when the phrase is at the first of the sentence ($\bar{X} = 2.0$) or at the end ($\bar{X} = 1.85$).

There is also a significant interaction ($F = 3.51$, $df =$

2/64, $p < .05$) between Concreteness and Phrase Position; the means are presented in Table 1-2 and Figure 1-1.

Insert Table 1-2 and Fig. 1-1 about here

Proportion of words recalled correctly. The results for this measure parallel closely those obtained for the previous measure. Concrete sentences had an average proportion of .67 words recalled correctly while abstract sentences had a mean proportion of .45 ($F = 19.3$, $df = 1/32$, $p < .01$). A lower proportion of words was recalled for the two-adjective case ($\bar{X} = .50$) than for either the one-adjective ($\bar{X} = .6$) or zero-adjective ($\bar{X} = .59$) instances ($F = 7.10$, $df = 2/64$, $p < .01$). A similar result obtains for Phrase Position; proportional recall when the phrase is absent ($\bar{X} = .64$) is better than when the adverbial phrase is first ($\bar{X} = .51$) or last ($\bar{X} = .52$) in the sentence ($F = 18.85$, $df = 2/64$, $p < .01$).

As well, there is a significant interaction ($F = 6.29$, $df = 2/64$, $p < .01$) between Concreteness and Phrase Position. The mean proportions for this interaction are presented in Table 1-3, and a graphical presentation is in Fig. 1-2.

Insert Table 1-3 and Fig. 1-2 about here

Forward deletion TEPs. There was a significant main effect for Number of Adjectives ($F = 10.63$, $df = 2/64$, $p < .01$). The zero-adjective case had a significantly lower mean forward deletion TEP (.011) than either the one-adjective condition (.020) or the two-adjective condition (.037).

The mean TEPs for the significant interaction between Phrase Position and Concreteness ($F = 3.53$, $df = 2/64$, $p < .05$) are presented in Table 1-4 and Figure 1-3.

Insert Table 1-4 and Fig. 1-3 about here

Backward deletion TEPs. The only significant effect was that of Number of Adjectives ($F = 6.88$, $df = 2/64$, $p < .01$). The zero-adjective condition had a significantly lower mean TEP (.010) than either the one-adjective case (.026) or the two-adjective case (.032). One possible explanation for this result is that Ss were learning some parts of the sentence (e.g., the subject phrase) in a right-to-left order. Thus, Ss would learn the subject noun first and the adjectives later. When there were no adjectives to be learned (the zero-adjective case), the mean backward deletion TEP for the sentence was lowered. Test of this

Table 1-2. Interaction of Phrase Position x Concreteness--Number of sentences entirely correct

		Sentence Type	
		Concrete	Abstract
Phrase Position	Beginning	2.61	1.39
	End	2.72	.98
	Absent	3.37	2.52

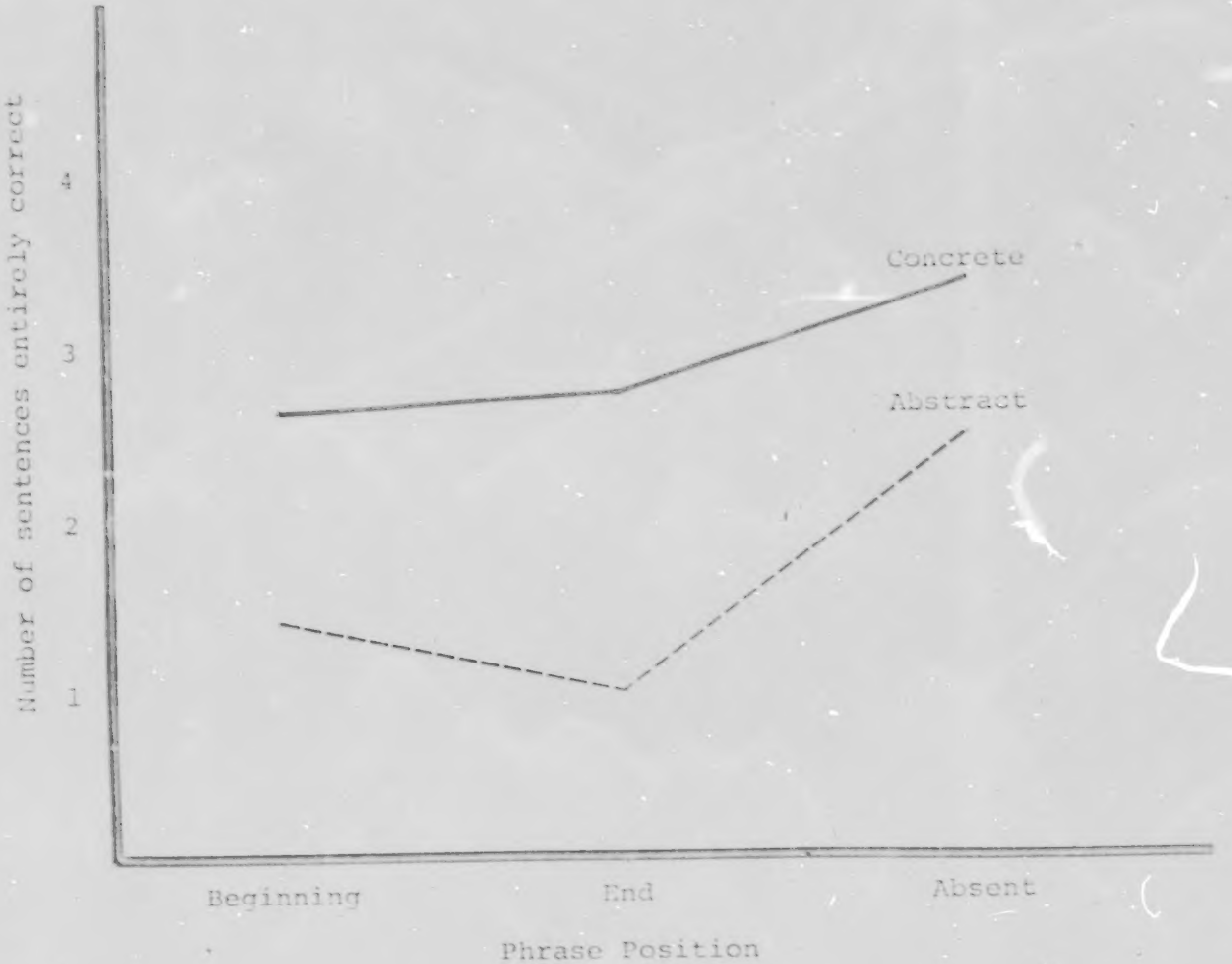

Fig. 1-1. Interaction of Phrase Position by Concreteness--Number of sentences recalled entirely correct.

Table 1-3. Mean proportion words recalled correctly--
Interaction of Phrase Position x Concreteness

		Sentence Type	
		Concrete	Abstract
Phrase Position	Beginning	.62	.42
	End	.68	.37
	Absent	.72	.57

Fig. 1-2. Interaction of Phase Position by Concreteness--Proportion of words recalled correctly.

Table 1-4. Interaction of Phrase Position x Concreteness--Mean forward deletion TEPs

		Sentence Type	
		Concrete	Abstract
Phrase Position	Beginning	.013	.022
	End	.017	.042
	Absent	.026	.016

Fig. 1-3. Interaction of Phrase Position by Concreteness--Mean forward deletion TEP's.

hypothesis must await review of the specific transition error patterns in a later section.

Forward commission TEPs. There was a significant effect for Number of Adjectives ($F = 5.07$, $df = 2/64$, $p < .01$). The two-adjective case had the highest mean TEP (.087) followed by the zero-adjective case (.071), then the one-adjective case (.056).

Backward commission TEPs. Only one significant result was obtained--an effect for Number of Adjectives ($F = 4.12$, $df = 2/64$, $p < .05$). The ordering was the same as for forward commission TEPs: the two-adjective case had a mean TEP of .066, followed by .053 for the zero-adjective case, and .041 for the one-adjective case.

Forward stop TEPs. Concrete sentences had a significantly lower mean TEP (.005) than abstract sentences (.035) ($F = 28.89$, $df = 1/32$, $p < .01$). The difference between these two means is especially striking if we convert these probabilities to mean frequencies. Stop errors occurred about once every three Ss for concrete sentences while each S learning abstract sentences made about two stop errors over his six trials.

Backward stop TEPs. The only significant result was for Phrase Position ($F = 4.68$, $df = 2/64$, $p < .05$). When the phrase was at the beginning, there was a significantly higher mean TEP (.017) than when the adverb phrase was absent or at the end of the sentence (both means were .006). One possible explanation is that Ss found the learning of sentences with adverb phrase at the beginning to be difficult because the sentence sounded stilted; they may have been recalling the sentence with the adverb phrase moved to the more natural position at the end of the sentence. This would result in higher backward stop TEPs for the condition when the phrase was presented at the beginning of the sentence. However, given that the adverb phrase (or part of it) was recalled at all, less than one half of one percent of the time did Ss recall it out of position.

Synonym errors. These are errors that were judged to involve replacement of a correct word by a synonymous one (e.g., replacing "diamond" by "jewel" in recall). Abstract sentences had a significantly higher proportion of words replaced by synonyms (.028) than concrete sentences (.018) ($F = 7.55$, $df = 1/32$, $p < .01$).

Omission errors. A significant effect for Number of Adjectives ($F = 6.56$, $df = 2/64$, $p < .01$) showed that there

were significantly fewer omission errors for both the zero-adjective (.36) and one-adjective cases (.35) than for the two-adjective case (.43). Adverb Phrase Position also had an effect ($F = 12.69$, $df = 2/64$, $p < .01$). When the phrase was absent, there were significantly fewer omission errors than when the phrase was at the beginning or end of the sentence (.30 vs .41 and .42). A significant effect for Concreteness ($F = 28.06$, $df = 1/32$, $p < .01$) revealed a large difference between concrete sentences (.26) and abstract sentences (.55).

As well, there was an interaction ($F = 5.01$, $df = 2/64$, $p < .01$) between Concreteness and Phrase Position. The mean proportions are presented in Table 1-5 and Figure 1-4.

Insert Table 1-5 and Figure 1-4 about here

Subject Phrase Analyses--Word Frames

Due to the importance attached to the variables manipulated in the subject phrase (association value of adjectives and number of adjectives), a three-factor analysis of variance design (Concreteness by Association Value by Number of Adjectives) was used to analyze individually each of the first 5 word frames on the following dependent measures: a) proportion of words correct, b) number of synonym errors, c) number of omission errors, d) number of commission errors.

In the analysis of errors in the individual word frames, the factor Number of Adjectives was treated as a two-level factor in the analyses of adjective errors because no adjectives occurred in the zero-adjective condition. As well, since there were no significant differences on proportion correct between adjective one and adjective two in the two-adjective condition, a mean value for the two was computed and used in the analyses.

Because of the large number of significant effects obtained, only the theoretically important ones will be reported. Most of the effects pertaining to "The" will be omitted.

Recall of the subject phrase noun. Two variables exhibited strong main effects in this analysis (and in the ones to follow). The concreteness of the sentence affected both the proportion of nouns recalled correctly ($F = 10.82$, $df = 1/32$, $p < .01$) and the number of omission errors ($F = 17.25$, $df = 1/32$, $p < .01$). Concrete nouns were recalled correctly 67% of the time while abstract nouns were recalled 48% of the time. As well, concrete nouns were omitted 24% of the time while abstract nouns were omitted 44% of the time.

The number of adjectives also affected both recall of

Table 1-5. Interaction of Concreteness x Phrase Position--
Mean proportion of words recalled as omission errors

		Sentence Type	
		Concrete	Abstract
Phrase Position	Beginning	.14	.41
	End	.13	.42
	Absent	.11	.30

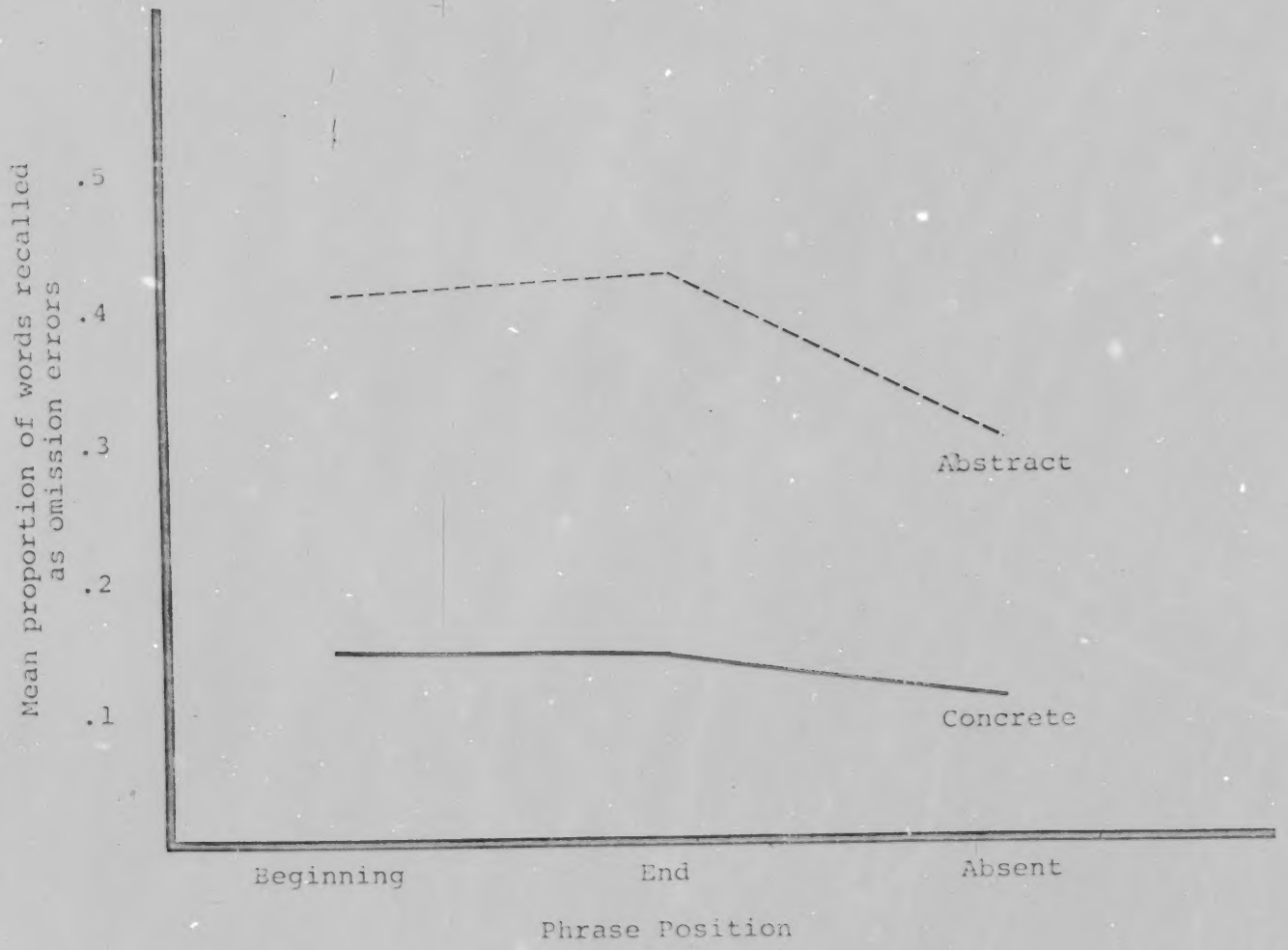

Fig. 1-4. Interaction of Phrase Position by Concreteness--Mean proportion of words recalled as omission errors.

the nouns (F =8.58, df = 2/64, $p < .01$) and the number of omission errors (F = 11.09, df = 2/64, $p < .01$). The noun was recalled less often in the two-adjective case (50%) than in either the zero-adjective or one-adjective cases (59% and 63%, respectively). When there were omission errors, they were more likely to occur in the two-adjective case (41%) than in the one-adjective (3%) or zero-adjective cases (28%).

Recall of the verb. Almost the identical pattern occurs with verb recall as does with the subject noun recall. Concreteness and Number of Adjectives yield the only significant effects, and they do so for proportion correct and number of omission errors. Concrete verbs were recalled correctly 63% of the time while abstract verbs only 40% of the time (F = 15.77, df = 1/32, $p < .01$). Concrete verbs were omitted 25% of the time while abstract verbs suffered this fate 50% of the time (F = 25.98, df = 1/32, $p < .01$).

In the two-adjective case, verbs were less frequently recalled (44%) than in either the one-adjective (56%) or zero-adjective (54%) cases (F = 9.83, df = 2/64, $p < .01$). As well, when two adjectives were present, there were more verb omission errors (4.4%) than when one (3.5%) or no adjectives (3.4%) were present (F = 10.22, df = 2/64, $p < .01$).

Recall of adjectives. When two adjectives were present, the average recall frequency was 50% while the one-adjective case yielded 59% recall (t = -2.31, df = 35, $p < .05$, two-tailed).

Between-Phrase--Within Phrase Transitional Error Probability Analyses

To examine what effect each variable had on unitization within the subject phrase, the mean forward TEP (deletion, commission, and stop) within the subject phrase was compared to the TEP for the subject noun-verb transition. Here the noun-verb transition is being treated as a reference datum. This analysis design is the same as the three-factor design used for the analysis of the word frames above, with the addition of Transition Type (within-phrase and between-phrase) as a repeated-measures factor.

Because the important question involves differential unitization between or within the phrases, only those significant results involving Transition Type will be reported.

Forward Stop TEPs. There was a significant effect for Transition Type (F = 34.57, df = 1/82, $p < .01$). The mean within-phrase TEP was .02 while it was .15 for the between-

phrase transition. The high TEP level for the between-phrase transition is important support for our contention that Ss were learning the phrases as relatively isolated units (chunks). There was also a significant interaction involving Concreteness ($F = 21.97$, $df = 1/32$, $p < .01$). Table 1-6 presents the relevant TEPs.

Insert Table 1-6 about here

As may be seen from Table 1-6, the significant interaction comes from the very high value for abstract between-phrase transitions. Given that the noun is recalled, there is a 27% probability that S will terminate his recall attempt at that point. The Ss are obviously treating the subject-predicate boundary as a major psychological boundary in their recall of abstract sentences.

Forward deletion TEPs. Again, there is a significant main effect for Transition Type ($F = 6.61$, $df = 1/32$, $p < .05$). Now the mean TEP is .05 for between and .11 for within the phrase. This result suggests that when Ss make a deletion error, it will much more commonly involve a word within the subject phrase than it will involve the verb. The locus of this effect is seen clearly in the significant interaction of Transition Type and Number of Adjectives ($F = 3.56$, $df = 2/64$, $p < .05$). Table 1-7 and Fig. 1-5 show that the value of the mean forward deletion TEPs is about constant for both the two-adjective case and the one-adjective case and plummets to zero with the zero-adjective situation.

Insert Table 1-7 and Fig. 1-5 about here

For the within-phrase situation, it is probable that what is being deleted is/are the adjective(s). When adjectives are present, they also exert an influence beyond the subject phrase; they affect the probability of deletion errors of the verb also. This effect is not predicted from Johnson's decoding-operation model; there is nothing in the model to suggest how characteristics of one chunk could affect recall of items in another chunk. Chunks may not be "opaque containers" after all. Another finding appropriate to mention at this time is a significant effect for Phrase Position on an analysis of the number of times the sentence was recalled completely wrong. Johnson's model predicts more omission errors for the sentence as a whole the longer the first chunk is. If the adverbial phrase is first in the sentence, the first chunk contains 3 items; if the phrase is either last or omitted, the first chunk still (on the average) contains 3 items (the subject phrase). Thus, there should be no clear differences because the lengths of

Table 1-6. Mean forward stop TEPs for interaction of Concreteness by Transition Type

	Transition Type	
	Within	Between
Concrete	.00	.03
Abstract	.04	.27

Table 1-7. Mean forward deletion TEPs for interaction of Transition Type and Number of Adjectives

		Number of Adjectives		
		2	1	0
Transition Type	Within	.174	.152	0
	Between	.075	.06	.015

Fig. 1-5. Interaction of Transition Type by Number of Adjectives--Mean forward deletion TRP's.

the first chunk are constant.

However, differences do occur. When the phrase is first, there are an average of 2.3 cases where the sentence is recalled entirely incorrectly. When the phrase is last, this figure is 1.7 and 1.4 when the phrase is omitted. This difference cannot be explained on sequence length either since the sentence lengths in the three cases above (in order)are: 9, 9, 6.

Obviously, something other than the size of the first chunk is affecting the number of times S has completely lost the sentence from memory. Probably, the disruption in familiar order of the sentence phrase elements is a great hindrance both to Ss storage and retrieval of the sentence. This is not to deny the validity of Johnson's model; the effect simply shows that there are other factors operating which can overpower the effect of the length of the first chunk on omission errors. Another prediction from Johnson's theory is that the number of chunks remaining to be recalled exerts a strong influence on stop TEPs at a chunk transition. However, no significant differences at the noun-verb transition were found as a function of number of remaining chunks.

All-or-None Learning

The last major analysis utilized the factors of Concreteness, Association Value, and Number of Adjectives to examine what effect the manipulations had on all-or-none recall of sentences. The dependent measure used for this analysis was a modified Z' clustering statistic (Flagg & Reynolds, 1972). Briefly, the clustering statistic is calculated as follows. For each recall attempt by each S a measure (O) is computed which describes the number of pairs of correctly recalled words in the recall string. For each possible string length, M, an estimate based on random placement of words during recall, is also computed. The difference between O and M divided by the predicted variance (V) gives an S-statistic which can be converted to a standard score (Z). To solve the problem of unequal ranges and variances with different string lengths, this score is divided by a constant to yield a Z' statistic (with a range of +1 to -1). The more positive this figure is, the more correctly ordered or clustered the recalled string is. If we omit from our analysis those cases where no recall attempt was made, the Z' statistic becomes a measure of all-or-none learning. An independence model of learning predicts essentially a random placement of items in a recall string. The more that items are recalled in the veridical order, the less an independence model holds. Of course, neither an independence nor all-or-none model has any specific predictions when recall is zero, hence this is omitted from our analyses.

None of the manipulated factors resulted in significant main effects. This is very discouraging given the strong theoretical grounds on which these factors were chosen.

The main findings of this study in relation to N. F. Johnson's decoding-operation model and previous research on sentence learning can be summarized as follows.

The finding of clear stop TEP spikes at major phrase boundaries verifies the argument that Ss treat phrases as chunks of information and that surface-structure grammatical form can function effectively as a segmentation procedure for the study of chunking of information.

The size of the initial chunk of information does not seem to be the main determinant of omission errors. Rather, the more disruptive factor is a manipulation involving placement of a phrase, normally found near the verb, at the beginning of the sentence. Although grammatically legitimate and semantically intelligible, the resulting sentence is very difficult for S to learn, store, or retrieve. As well, it was found that the number of chunks remaining to be recalled did not affect stop TEPs at the noun-verb transition.

The main questions which generated this study remain unanswered--those related to the variables which affect all-or-none learning. No significant differences were obtained with any of the variables employed. It is possible that the dependent measure employed (Z' statistic) is either unsuited or insensitive in this situation. Attempts will be made in the future to find more promising measures.

The instances where clear significant effects were found are numerous nonetheless. The concreteness of the stimulus sentences almost never failed to be significant on any of our several measures of recall performance. This is in line with the reported potency of this variable (e.g., Paivio, 1971) under many experimental paradigms. It is clearly seen in the present study that concreteness also has large effects in paired-associate sentence learning. In the studies to follow this potency will again be demonstrated many times; thus, a discussion of the possible theoretical reasons for this will be postponed until the end of the report.

Given the encouraging findings regarding phrase structure as a chunking/segmentation cue, it was felt that a developmental study was in order to attempt to assess the beginnings of this strategy in childhood. The next experiment explores this and several related questions.

Experiment 2

This study was designed to explore some of the possible developmental processes which would lead to the types of chunking performance uncovered in Exp. 1. The literature in this area is scant. Only one study (Matthews, 1965) has explored the developmental trends in chunking as treated by Johnson's model. This study (an unpublished M. A. thesis) examined the patterns of sentence TEPs of children in grades one through four using a paired-associate task. The main dependent measure was the correlation between obtained TEP pattern and the predicted pattern based on Johnson's decoding-operation model. The mean correlation showed an abrupt jump between seven and eight years of age. This result is consistent with previous word association data suggesting a shift from syntagmatic to paradigmatic response styles at this period. The results can be interpreted in two (not necessarily independent) ways. First, the pattern may represent an increase in the children's functional knowledge of parts of speech. It may also indicate a change in the functional knowledge of grammar.

In an attempt to explore the consistency of this pattern we decided to ask children to perform a sentence-learning task (as Matthews, 1965, did) and also to take a standardized word-association test. We hoped that administration of both measures on the same children would clear up some of the conceptual haziness deriving from Matthews' use of word association norms only.

Initially, we attempted a paired-associate procedure as was done by Matthews but soon uncovered a surprising finding. Although the objectively-scored measures of word recall agreed well with the findings reported by Matthews, it was clear that the results were in some sense spurious. Take, for example, the performance of children in the first grade. Over five trials of paired-associate learning with six number-sentence pairs per trial, children were showing little evidence of correct recall until the third or fourth trials. This was not due, however, to lack of recall or faulty recall within a sentence but rather to many instances of complete (correct) recall of a sentence to the wrong number cue. We felt that TEPs based on correct recall to inappropriate cues would possibly be misleading. It seemed evident that even the very young children were learning the sentences fairly quickly but were simply failing at the number-sentence linkage. For this reason we modified our learning task to that of a probe-recall operation in the hopes that it would give a purer measure of the integration and chunking strategies used by the children.

Method

Subjects. Twenty students from each of grades 1-4 from the Wilder (Vt.) Elementary School and twenty volunteers from Introductory Psychology at Dartmouth College served as Ss.

General design. The task was a modified probe-recall procedure where S was presented with a block of six sentences for a presentation trial and then given one of the phrases of one of the sentences as a cue for recall of the sentence. All six sentences were tested for recall on each trial; there were five trials with order of administration of both the cue phrase and the sentences varying from trial to trial. After S had responded on each presentation, the full correct sentence was read to him. There were two types of sentence materials (concrete and abstract) and three possible cues (subject phrase, modifier phrase, or verb phrase). The general design thus employed the factors of Grade (one through four plus adults), Concreteness (concrete or abstract sentences), and Probe (three possible phrase cues). The first two factors are independent; ten Ss were tested in each cell of the 5 x 2 design. The last factor is a repeated-measures variable.

Materials. All sentences used were of the general form "The wall around the grey castle was high." The three possible cue phrases are: 1) The wall, 2) around the grey castle, and 3) was high. Six abstract and six concrete sentences of the above form were generated by E from a pool of AA count words (Thorndike & Lorge, 1944). These sentences were independently evaluated by two colleagues for ease of understanding, meaningfulness, and comparability of difficulty. In order to make the sentences as representative as possible, a second form of each sentence was constructed where the only change was the substitution of another AA adjective in the modifier phrase. These two parallel forms of the sentences were administered to equal-sized groups of Ss; since there is no theoretical issue at stake, no analyses were performed on the alternate versions, and hereafter this manipulation will be ignored.

The word-association stimuli used were the first twenty-five stimulus words of the Kent-Rosanoff list with the exception of the word "mutton" which many of the younger children did not understand.

Procedure. Testing of the elementary school children was individual; all sentences were presented orally by a young female experimenter with good rapport with the children. The sessions were tape-recorded and later transcribed.

After instructions using a practice sample, E read each of six sentences twice; S was asked to repeat each sentence.

Trial 1 began with E giving S a probe phrase. The probe was presented as a "clue" or a "part of the sentence," and S was asked to remember or "guess" the whole sentence. After a response was given by S, E corrected any mistakes by reading the full correct sentence and asking S to repeat the correct answer. If S had responded correctly, E gave verbal approval and repeated the correct answer. This continued for five trials with the order of presentation of sentences varying from trial to trial but consistent for all Ss. Children received either concrete or abstract sentences. After the sentence-recall part of the study, children were tested on a 24-item word-association test with standard instructions ("Tell me the first word that you think of, etc."). At the conclusion of all testing each child received a small reward (toy, charm, etc.).

The procedure was essentially the same for the twenty adult Ss with two exceptions: Ss were tested in two groups of 10 each, and responding was written, not oral. The same E that tested the children administered the task to the adults.

Results and Discussion

Preliminary analysis of the data indicated that the five groups (grades 1-4 and adults) differed strikingly in the extent to which they used the cue phrase in their recall attempts ($F = 14.84$, $df = 4/90$, $p < .001$). While adults almost always responded with the cue phrase (99%) even when this was all they could answer, children in grade 1 used this strategy only 78% of the time. Thus, it was decided to omit the cue words from all subsequent analyses.

Proportion of words recalled correctly

There was a striking main effect for Concreteness ($F = 162.25$, $df = 1/90$, $p < .001$). For concrete sentences recall was 87% while it was only 58% for abstract sentences.

There was also a main effect for Probe ($F = 10.56$, $df = 2/180$, $p < .001$). The subject phrase (74%) and the modifier phrase (73%) functioned as better cues than the verb phrase (70%). This result may be partly specific to the sentences used in the present study since the verb phrases were more general in meaning and could meaningfully be used in more than one sentence. There was a significant effect ($F = 14.47$, $df = 4/90$, $p < .001$) for Grade. This effect is overshadowed by the significant interaction with Concreteness ($F = 10.40$, $df = 4/90$, $p < .001$). This finding is presented in Table 2-1 and Fig. 2-1.

Insert Table 2-1 and Fig. 2-1 about here

As may be seen from Fig. 2-1, there is almost no difference

Table 2-1. Mean proportion of words recalled correctly for interaction of Grade by Concreteness

		Concrete	Abstract
Grade	1	.85	.34
	2	.85	.52
	3	.92	.54
	4	.86	.64
	Adults	.90	.84

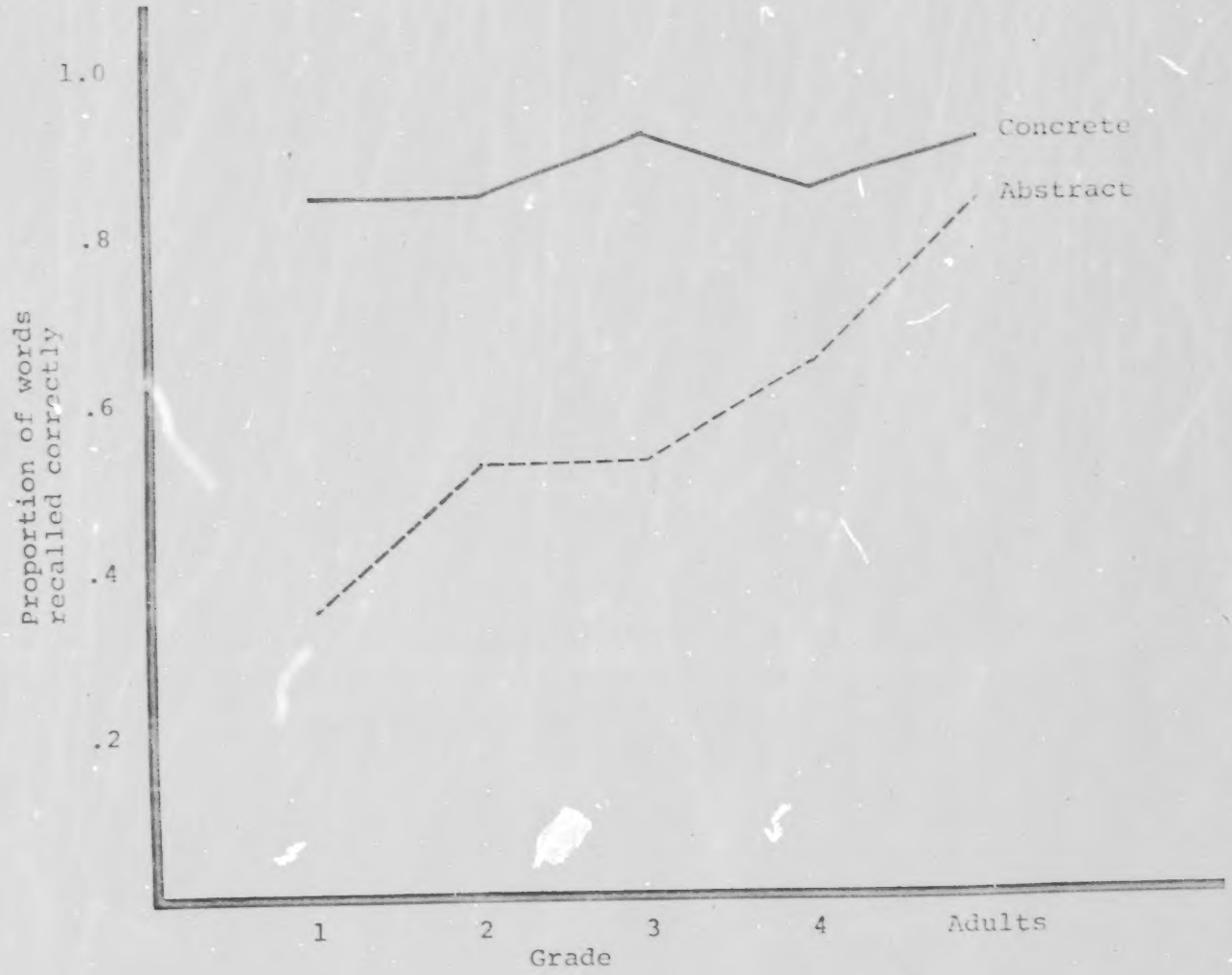

Fig. 2-1. Interaction of Concreteness by Grade--Proportion of words recalled correctly.

on recall of the words in the concrete sentences. Although recall is not perfect, it is high for all age groups. There is no superiority for adults over children. As a matter of fact, performance by the adults is actually inferior to that of children in grade 3. For abstract sentences there is steady improvement in recall with age.

These results raise an important point. It has long been a central feature of many developmental theories of cognition that the use of ikonic imagery is a prime representational technique employed by children. It is only with maturation that this ability becomes overshadowed or replaced by the use of verbal memory. The present finding demonstrates clearly that given the appropriate materials and experimental technique, the use of imagery is probably not lost by adults and in fact is present as an effective mnemonic mechanism in very young children. It is a moot point as to how far back on the chronological scale we would have to proceed in our experimental design before we found a decline in performance with concrete materials.

Paraphrase Scoring

To explore the possibility that the results obtained above for verbatim recall are in fact representative of the memory storage abilities of the children, a further analysis was performed. It may be that the type of mental representation employed by the children is essentially different from that of adults, but through the use of verbatim scoring it appears similar for concrete materials. That is, the adults may greatly surpass the children's performance when we use a less restrictive measure of recall. Hence, we decided to score the sentence recalls not for verbatim errors but rather for failure to recall the essence or core of the presented sentence. Two judges, naive as to the purpose of the experiment, independently scored all recall attempts. They disagreed less than 1% of the time as to what constituted a failure to recall the essence of the sentence. These paraphrase errors were then analyzed in the type of analysis of variance design used above.

The results almost completely match those for verbatim scoring of number of words recalled correctly. The most interesting result is the significant interaction of Concreteness by Grade ($F = 7.34$, $df = 4/90$, $p < .001$). Here the pattern is identical to that observed above. For concrete sentences grades 1, 2, and 4 are essentially equal in performance, adults and grade 3 are the best. For abstract sentences there is a steady increase in performance from grade 1 to adults.

This result verifies that the finding for number of words recalled correctly was not artifactual.

Clustering

Not only are we interested in the number of words the children recall but also whether they are recalled in their appropriate places. The clustering measure (Z') described in Experiment 1 was employed as a dependent measure here. Besides a significant main effect for Grade ($F = 15.09$, $df = 4/90$, $p < .001$), there was also a significant interaction between this factor and Concreteness ($F = 5.94$, $df = 4/90$, $p < .001$).

The results are identical to those obtained above: for concrete sentences grades 1, 2, and 4 are surpassed by grade 3 and the adults. For abstract sentences there is a steady increase in clustering from grade 1 to the adults.

Word-Association Analyses

The 24 word-association responses from each child were scored on a number of categories. These included the standard categories of paradigmatic, syntagmatic, klang, and blanks. As well, a fifth scoring category was devised: every response a child made that exactly matched the primary adult response for that stimulus was scored as "adult-like." Analyses of variance were performed on all these response measures.

Two of the dependent measures are of special theoretical importance to us: adult-like and paradigmatic. It is widely accepted that the shift to paradigmatic responding is characteristic of maturation processes. Both these measures then may be assumed to be indicators of the degree to which the children perceive form-class relations as adults do on a word-association test.

The analyses for both adult-like and paradigmatic responses yielded significant effects for Grade ($F = 30.99$, $df = 3/76$, $p < .001$ for adult-like and $F = 46.63$, $df = 3/76$, $p < .001$ for paradigmatic). It had been hoped that the pattern of means would show some resemblance to the sentence recall patterns displayed by the children. However, there was no clear indication that the grade 3 children displayed any differential form-class knowledge. For adult-like responses grade 1 had a mean of 3.5, grade 2 had 9.1, grade 3 had 8.8 while grade 4 had 11.8. The same ordering occurs for the paradigmatic category. It thus appears that the pattern of recall displayed by the children on the sentence probe task cannot be explained on the basis of differential form-class knowledge.

More likely, the effects noted earlier can be traced to the growth of the memory representational systems rather than to retrieval strategies at test time. The data strongly suggest that the children differ from the adults not in the degree to which they unitize and recode sentential information but rather in the manner in which such information is stored. To substantiate this idea,

correlations between each child's average TEP pattern on the sentences were correlated with the average adult TEP pattern for the same types of sentences. There were no significant differences across grades for the average degree of similarity to the adults' pattern of recall. However, as we saw earlier, there were consistent and striking differences across age levels for number of words recalled regardless of the method of scoring. These results reinforce our contention that the maturational sequence observed in the children under study reflects not a different means of chunking and organizing sentential information but rather the growth of a second storage system for the storage of abstract information.

Experiment 3

It would appear from previous research (Begg and Paivio, 1969; Tulving and Patkau, 1962) that the size and number of chunks of information is determined primarily by the nature of the stimulus materials. The specific qualities of the stimuli which affect chunk formation are still vague. As well, the conditions necessary for the formation of hierarchies of chunks ("super-chunks") are still largely unknown. One important factor which does seem to reliably affect size and number of chunks formed is memory load imposed on the Ss. The assumption that the maximum number of chunks that may be stored is a fixed value (7±2 for Miller, 1956a; 5 for Simon, 1970) leads the way to this experiment.

To the extent that formation of super-chunks is dependent on the memory load imposed on S, embedded clauses should be more likely to be subunits of verb chunks as memory load is increased. With low memory load conditions, embedded clauses could stand as chunks on their own. It is also expected that concreteness of the sentences should facilitate formation of super-chunks and also result in confusion of placement of auxiliary clauses if Ss do in fact unitize concrete sentences more readily than they do abstract ones (memory for abstract sentences is assumed to be tied more closely to the left-to-right order of events in the sentence).

Method

Subjects. A total of 89 students from the Introductory Psychology course at Dartmouth College served as Ss; 39 of these helped in preparation of the stimulus sentences, and 50 Ss were tested in the experiment proper.

Design. The experiment was a two-component memory task (cf. Savin & Perchonock, 1965); Ss were presented with digit strings followed by sentences and were instructed to recall both the digits and the sentences. It was emphasized in the instructions that it was more important to remember the digits but to try to recall both if possible.

The experiment was run as a factorial design with two independent factors and one repeated measures factor. The two independent factors were Concreteness of the sentences (concrete and abstract sentences were used) and Digit Length (the digit strings varied in length from one to five numbers). The repeated measures factor was Clause Position (the subordinate clause could occur in the middle of the sentence or at the end of the sentence).

The design was thus a 2 x 5 x 2 factorial design with one repeated measure; five Ss were tested in each of the

ten cells formed by the two independent factors.

Materials. The sentences all were of the form: The national election, since the results were clear, indicated security for the country. The 40 concrete and 40 abstract sentences used in the study were generated in the following manner. A group of 39 students from Introductory Psychology created a corpus of abstract sentences of the above form. This pool was then evaluated by the experimenters; any sentence which did not fit the sentence form, did not make sense, or was questioned on abstractness, was discarded. Of the 60 sentences which remained, 40 were chosen randomly to be used as stimulus sentences. The Thorndike-Lorge frequencies for the adjectives, nouns, and verbs in these 40 abstract sentences were compiled, and 40 concrete sentences matched word-for-word on Thorndike-Lorge value within a range of ±5 were generated. All concrete sentences were independently approved by three judges on the criteria of concreteness and meaningfulness. Two versions of each concrete and abstract sentence were constructed by simply placing the subordinate clause in the middle or at the end of the sentence.

Procedure. Testing of Ss took place at a computer teletype connected to the Dartmouth Time-Sharing System; Ss were prevented from seeing what was being printed by a small screen taped to the teletype. When an entire line had been printed, this was advanced into view so that S could see it all at once. After S was seated at the teletype and had read the instructions, he was presented with 40 learning trials involving the digit strings and sentences.

On each trial, a digit string was in view for 5 sec, followed by a sentence for 10 sec. After this study period, S was presented with a letter of the alphabet and was instructed to recite the alphabet backwards from that point (for 4 sec). At this point S was asked to recall the digit string and then the sentence. No feedback was given to S concerning the correctness of his answers. This testing procedure continued until all 40 different digit string-sentence groups had been presented; S then completed a questionnaire to assess mnemonic strategies.

Results and Discussion

The initial analyses in this study utilized the three factors of Concreteness of the sentence (concrete or abstract sentences), Digit Length (one to five digits to be recalled), and Clause Position; the last factor is a repeated-measures factor.

The first analyses showed a disturbing finding for the ten Ss in the five-digit group. They displayed a clear

inferiority in terms of their accuracy of recall of the digits (75% recalled correctly vs 87% for the average recall performance of the other four groups); at the same time they displayed a much higher score for proportion of words in the sentence recalled correctly (80% vs 70% for the other four conditions). Since the role of the digits was to occupy a fixed amount of memory capacity (within each group), it initially appeared that this function was not being performed in the five-digit group.

It seemed that there may have been an attentional shift on the part of some or all of the Ss in the five-digit group such that they were not paying as much attention to the digit-recall part of the paradigm as they were to the sentences. Several attempts were made to correct for this possibility. The two Ss in each of the five digit-length groups who manifested poorest digit recall were deleted; then, the analyses were re-performed, omitting any trials in which recall of the digit strings was imperfect. The same findings emerged for the five-digit group: poorer recall on digits, better recall of the words of the sentence.

A tentative decision was made at this time to omit the five-digit group from the design since it appeared that Ss in this group were somehow approaching the experimental task quite differently from Ss in the other four groups. At this point, the post-experimental questionnaires of all Ss were examined in an attempt to find a clue for this differential performance. Over 60% of the Ss in the five-digit group reported consistent use of segmentation and chunking procedures to aid in recall of the sentences (cf. Johnson, 1969). It may be, then, that the five-digit condition, while depressing recall of the digits, facilitates recall of the sentences by driving Ss to a shift in recall strategy for the sentences. Regardless of the instructions, Ss may try to divide their attention equally between digits and the sentence--when the digit string becomes hard to handle due to length, Ss try to maintain sentence recall performance by adopting a new chunking technique. There is some suggestion that attentional shift is not to blame for the recall pattern in the five-digit group since the correlation between number of errors on digit recall and number of words recalled in the sentences is only .17. What then is the cause of this recall pattern? A partial explanation emerges if we compare the proportion of words recalled correctly (in the five-digit group) when the digit-string was recalled incorrectly to the number of words recalled when the digit string was correctly recalled. If an attentional shift was operating, then we should predict that more words would be recalled when the digits were misrecalled. This is not the case. When the digits were misrecalled, 76% of the words were recalled; when the digits were recalled perfectly, 81% of the words were recalled (t = 1.79, df = 9, p = .05, one-

tailed). There appear to be two processes going on. First, on some presentations there is a lapse of concentration, and Ss misrecall both the digits and parts of the sentence. On these instances Ss' recall of the sentences approximates the values found for the other four digit-groups (about 70%). On other occasions it may be that Ss adopt a highly effective (but possibly attention-straining) chunking strategy for both the digits and the sentences. This leads to perfect recall of the digits and good (81%) recall of the words of the sentence. Furakawa, Suydam, and Miller (1969) have demonstrated how a similar effect may be obtained by giving explicit grouping instructions to Ss.

Since the hypothesis of an attentional shift in the five-digit condition cannot be completely ruled out, it was decided to leave the five-digit group in the design but to analyze only those recall attempts (for all five groups) where the digit string was recalled correctly. All analyses must now be considered in the light of the fact that the five-digit group appears to be using different recall strategies than the other four groups.

Proportion of Words Recalled Correctly

There were two significant main effects. The words of concrete sentences were recalled better (84% recalled correctly) than the words of abstract sentences (61% recalled) ($F = 45.14$, $df = 1/40$, $p < .001$). As well, the length of the digit string had an effect ($F = 3.78$, $df = 4/40$, $p < .02$). The proportions recalled correctly, going from digit length of one to five are: .74, .73, .76, .61, .81.

Proportion of Sentences Recalled Perfectly

Concrete sentences were recalled perfectly 32% of the time while abstract sentences were recalled perfectly only 8% of the time ($F = 39.4$, $df = 1/40$, $p < .001$).

There was also a significant interaction between Digit Length and Clause Position ($F = 2.72$, $df = 4/40$, $p < .05$). However, this interaction is overshadowed by the significant three-way interaction involving these two factors and Concreteness ($F = 2.68$, $df = 4/40$, $p < .05$). This interaction is presented in Table 3-1 and Fig. 3-1.

Insert Table 3-1 and Fig. 3-1 about here

There are several very interesting effects evident from inspection of Fig. 3-1. Look first at the abstract sentence curves. First, there is no essential difference in performance as a function of placement of the subordinate clause; the two curves overlap almost completely. Second, the ease of recall is ordered thus: five-digit group, then three-digit group, then the other three groups.

Table 3-1. Interaction of Digit Length, Clause Position, and Concreteness for proportion of sentences recalled perfectly

		Concrete		Abstract	
		Clause in middle of sentence	Clause at end of sentence	Clause in middle of sentence	Clause at end of sentence
Digit Length	1	.49	.36	.06	.02
	2	.32	.36	.02	.04
	3	.21	.28	.13	.14
	4	.12	.28	.04	.02
	5	.43	.32	.17	.19

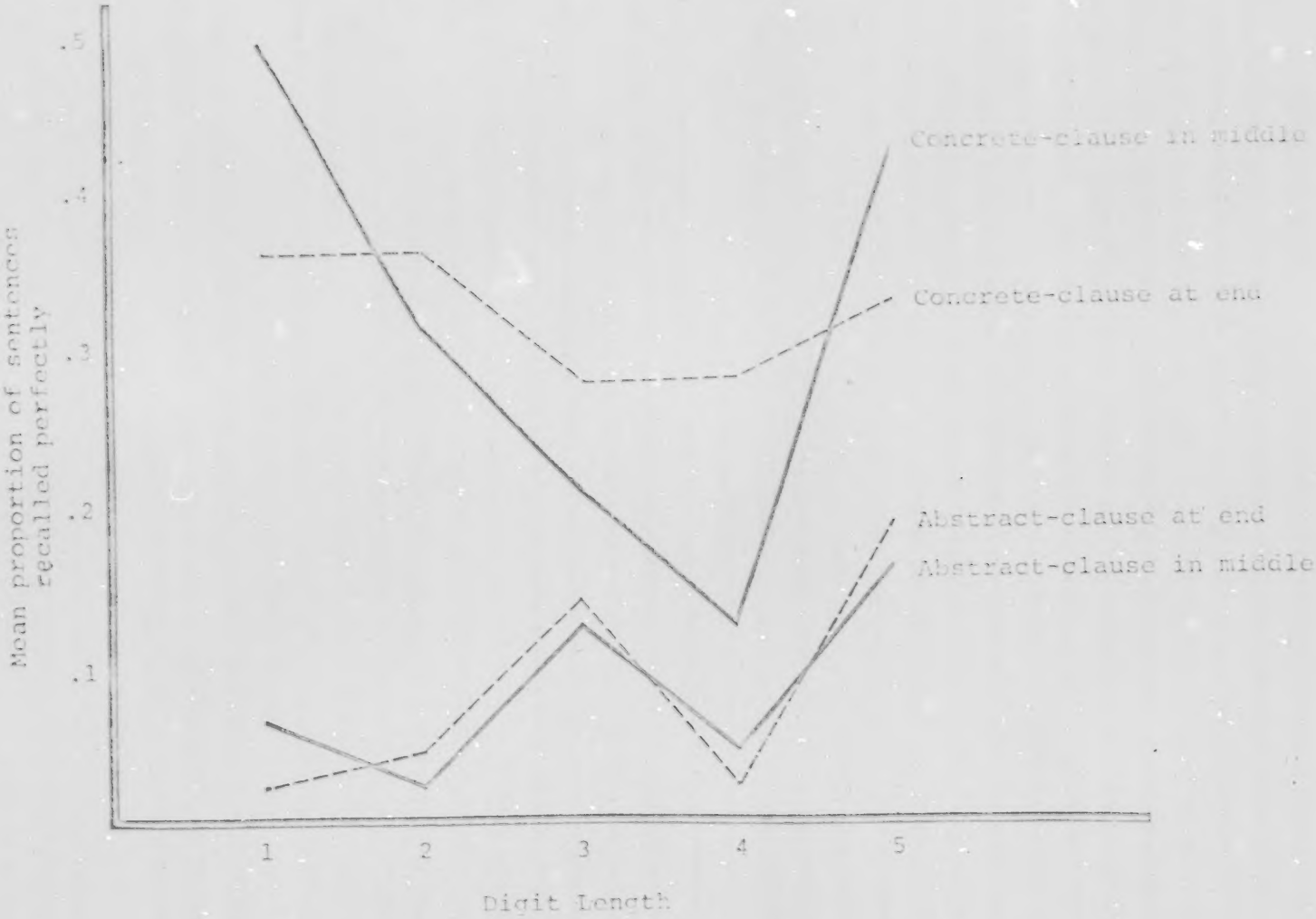

Fig. 3-1. Interaction of Digit Length by Clause Position by Concreteness-- Proportion of sentences recalled perfectly.

Quite a different picture emerges when we examine the curves for concrete sentences. No longer is there a clear similarity of performance for the two positions of clause placement. When the clause is at the end of the sentence, recall is almost uniform for the five-digit groups. However, when the clause is in the middle of the sentence, the number of sentences recalled perfectly correctly steadily declines from the one-digit group to the four-digit group. The five-digit group is much better, almost as good as the one-digit group.

It would appear that the insertion of the subordinate clause between the subject noun and the main verb has a substantial effect for concrete sentences. One possibility is that this manipulation effectively destroys the psychological unity of the whole-sentence idea (in concrete sentences) and thus interferes with recall (or storage). No such effect operates with abstract sentences since it is assumed that the mode of memory representation is more closely related to word-to-word relations.

It must still be explained why this hypothesized disruptive effect of subordinate clause placement in concrete sentences interacts with number of digits to be remembered. In line with the earlier discussion the five-digit group is probably using an entirely different form of processing strategy. But what explains the steady decrease in recall performance from the one-digit case to the four-digit case?

A possible explanation emerges if we assume that the length of the digit-string is having the desired effect of loading up memory, at least in the cases where the string length is less than five. If the placement of the subordinate clause in mid-sentence is disruptive (in concrete sentences) to the formation of global images, S may elect to store the clause in such a way as to minimize this effect. It may be stored as a subsidiary image or tag, but it must be invested with the appropriate order information for it to be recalled in the proper place in the sentence. This order information is also a component of memory load in a fixed-capacity system. Thus, increasing the digit length will increase the storage requirements for the sequence even more.

Boakes and Lodwick (1971) have suggested a different locus for this effect. The increase in digit-length may have its effect not on the fixed-capacity system as we assume, but rather the effect may be due to the increasing difficulty of recalling digit strings of greater length. It is the interfering effect of the recall procedure itself, they suggest, that decreases recall of the sentence. Either way the effect in the present study is a relatively weak, though interesting, one. The interaction did not occur when we analyzed proportion of words correct. As well, it is

difficult to specify the exact nature of the decline in performance. It is not (as we shall see later) a simple loss of the order information (re: the subordinate clause) since the effect does not appear when we analyze for errors of clause placement.

Chunking

There were three main questions related to chunking asked at the beginning of this study: 1) As memory load increases, do subordinate clauses become subunits of verb chunks? 2) Does concreteness of sentences facilitate the formation of super-chunks (unitization)? 3) Does concreteness result in confusion of placement of subordinate clauses?

The answer to 3) is an unequivocal "No." There was no significant difference between concrete and abstract sentences for confusion in placement of the subordinate clause ($F = .83$, $df = 1/40$).

To answer questions 1) and 2) it is necessary to examine both the Z' clustering data and the TEP data. The Z' analysis shows that concrete sentences are better clustered ($Z' = .79$) than abstract sentences ($Z' = .60$) ($F = 22.88$, $df = 1/40$, $p < .001$). This gives a qualified "Yes" to 2) in that it appears that concrete sentences are recalled in a more veridical order than the abstract sentences. However, to verify this answer for 2) and to find the answer for 1), it is necessary to examine the TEP patterns in recall.

In the sentence, "The national election, since the results were clear, indicated security for the country," two major grammatical boundaries are followed at transitions 3 to 8 by the words "indicated" and "since." This is also true when the subordinate clause is placed at the end of the sentence. First, the mean TEP for each of the twelve transitions in the sentences was calculated for each S for each of the two possible clause placement conditions. If we form a ratio of the mean TEPs at these two major boundaries, divided by the mean TEPs of the other ten transitions, the resulting measure describes the degree to which S is treating the three main grammatical sections as independent units. The higher this ratio, the less the sentence is being treated as a unitized whole; the lower this ratio, the more S is treating a between-phrase or between-clause transition like a within-phrase transition. This ratio turns out to be 2.8:1 on the average. However, an analysis of variance using the factors employed above failed to yield any significant effects.

It appears then that an increase in memory load (at least with the technique employed in this study) does not facilitate the formation of superchunks (as assessed by the TEP patterns at clause and phrase boundaries). It also

appears that the concreteness of the sentences does not have this effect either although concreteness does have a strong effect on recall in correct order for the sentences (and number of words recalled). The question of the role of concreteness in determining ordered recall will be examined in more detail in the next experiment.

Experiment 4

A last important question which has generated little research is "How are chunks kept ordered in memory, and what determines order of production in recall?" Is it some kind of left-to-right processing or primacy effect (cf. Johnson-Laird, 1968)? Since the verb is typically the poorest recalled element in "The subject verbed the object" sentences, it might even be a serial effect. However, this left-to-right emission might be superimposed on the sentence elements after recall or retrieval from memory. Johnson's (1972) model suggests two ways in which order information is preserved. First, by the very nature of the decoding steps, only certain elements are available for read-out from memory at any given time. Second, he hypothesizes that each element is tagged for order information as regards its position in the chunk and sequence of chunks.

This experiment is designed to answer two questions. There is the possibility that the left-to-right emission of sentence elements may not, in fact, be the order in which these elements were recovered from memory. This prediction, of course, is exactly opposite to that generated by Johnson's (1970) decoding-operation model. To explore this possibility, interchangeable subject and object phrases will be used in a recognition task. To the extent that material is not recalled in a left-to-right order, phrases in the subject position and object position should be often interchanged upon recall. This would imply that the order typically observed in emission is due to the organizing effects of grammar and not to any ordering during storage. We should expect any confusion to disappear when a valuable cue (the "by" element in the passive voice) is added. Although the present experiment is not designed as a direct test of predictions from Johnson's model regarding order information, nonetheless the results may have some important consequences for the model.

The second problem considered here involves the question of what is actually stored before the recognition trial. By manipulating the ways in which parts of distractor phrases are constructed, further information concerning how Ss store concrete and abstract words (semantically or by surface appearance) will be obtained.

There is a possibility that chunks may be "available" in memory but not "accessible" to recall unless some retrieval cue is present (Wood, 1969). Is one word of the chunk the best cue? Is this one word a noun (in a noun phrase)? If this is the case, it would be an example of "redintegrative memory" (Horowitz & Prytulak, 1969). Horowitz and Prytulak presented convincing evidence that nouns have some special importance in memory for sentences. In phrases the noun is the better cue than an adjective

(Lockhart, 1969); as well, in free recall the noun is better recalled than the adjective (Horowitz & Prytulak, 1969). They feel that the noun is the "most salient" fragment. Note, however, that most stimuli used were concrete, and Begg and Paivio have offered a more appealing explanation of the ease of recall of nouns in concrete sentences.

Anderson (1963) presented data on the ease of recall in entire sentences of the type "The subject verbed the object." He found that subject nouns were recalled best, then object nouns, then verbs. Some additional data on this issue come from Reynolds (1970). In this study Ss had to transform active-affirmative, active-negative, passive-affirmative, and passive-negative sentences. The Ss seemed to rely heavily on what Wright (1968) called key words (verb, object, and especially the subject). The propensity to do this increased as short-term memory load was increased. Horowitz and Prytulak have tied these and other findings together by showing that order of recall of an item and efficiency of that item as a cue are perfectly correlated.

Some additional information on the cue-value of nouns comes from a study by Roberts (1968). He found that bidirectional Subject-Object associational linkages aided sentence recall. When one chunk was correctly recalled, the noun in the phrase could act as a mediated cue for the other phrase.

Dolinsky and Michael (1969) have offered a different explanation for findings concerning the efficacy of the noun as cue. They argue that the noun has special importance in word phrases because of its role as a post-integration mechanism. Uhlenbeck's (1963) "principle of sustained memory" suggested that the elements in a phrase such as articles and adjectives may remain unconnected until a later element (e.g., the head noun) organizes them. Some empirical support for this argument was offered by Dolinsky and Michael (1969) and by Rommetveit and Turner (1967).

Quite a few factors seem important suspects as efficient cues for chunk retrieval. Nouns seem clearly to have a special role as mediators for phrase recall. As well, nouns appear to play different roles for concrete and abstract sentences. The effective relationship between nouns and their adjectives may be affected by several factors among which is the location of the noun phrase (subject or object). A similar set of relationships may occur with verbs as cues. Verbs may well function differently (as cues) depending on the voice of the sentence.

Method

Subjects. A total of 80 Ss participated in the study; 32 Ss

served in the experiment proper while 48 Ss helped generate stimulus sentences.

Design. The experiment used a recognition memory paradigm where five sentences were presented in a block for S to study; after an interpolated activity, a phrase (either unchanged or altered) from one of the five sentences was presented to S who had to signify if he had seen this phrase in the sample of five sentences. If S felt he had seen the phrase, he was asked to recall as much of that sentence as possible. The experiment was run as a factorial design with two independent factors and three repeated-measures factors. The two independent factors were Concreteness (concrete or abstract sentences), and Voice (sentences were either active or passive). Eight Ss were tested in each of the four cells formed by the two independent factors.

There were three repeated-measures factors; the last two of them define the types of alteration possible for changed sentences. The Cue Phrase Presented could be either the subject phrase, verb phrase, or object phrase; the Part of Speech Changed could be either the modifier or head in the phrase; the Type of Change could be a change to a synonym or to an unrelated word.

Materials. All sentences had the general form Article, Adjective, Noun, Adverb, Verb, Article, Adjective, Noun (e.g., The deep anger quickly caused the noisy outburst). All sentences were reversible; that is, the subject and object phrases could be interchanged without creating a nonsense sentence.

A group of student volunteers generated a pool of over 300 abstract sentences of the form described above. These were examined by E, and any sentences not meaningfully reversible (or otherwise inappropriate) were discarded. A few more random discards pared the pool to 240 sentences. From this pool, 48 sentences were randomly chosen as test sentences; the remaining 192 sentences were designated as fillers. To control for any semantic biases generated during the construction of the sentences, every second sentence had the subject phrase and object phrase reversed.

A computer program selected nouns, verbs, adjectives, and adverbs from a pool of concrete words to generate sentence frames that matched the 48 abstract sentences word-for-word on Thorndike-Lorge values (with a tolerance of ±5 on frequency). The 192 concrete filler sentences were gathered from the same subjects who produced all the abstract sentences.

After the concrete and abstract sentences had been constructed, each sentence was transformed into the passive form and checked for meaningfulness. Half the Ss were presented with the active form of each sentence, half with

the passive form.

Altered versions of 24 of the test sentences of each kind (concrete and abstract) were constructed in the following manner. Eight alternates had a change performed in the subject phrase, eight in the verb phrase, and eight in the object phrase, (Cue Phrase Presented factor). Within each group of eight alternate sentences, four had the modifier (adjective or adverb) changed, and four had the head (noun or verb) altered (Part of Speech Changed factor). Within each of these four sentences, two of the changes involved replacement by synonyms; two of the changes involved extraneous (neither synonymous nor antonymous) words. All replacement words were matched to the original words as closely as possible on Thorndike-Lorge frequency.

There are thus 12 possible changes that could be made in one of the test sentences. Since there were two exemplars of each of these 12 types constructed, and each exemplar had an unaltered counterpart sentence, one half of the unaltered test sentences were treated as if they were altered ("new") sentences for purposes of controlling any inadvertent biases in the construction of the altered sentences. Thus, one half of the original sentences were moved to the altered pool, and their altered counterparts were treated as if they were the original sentences.

Procedure. Testing of Ss took place at a computer teletype connected to the Dartmouth Time-Sharing System; after S was seated at the teletype and had read the instructions, a block of five sentences came into view. One of these five was the target sentence; four were fillers.[4] This block was exposed for twenty seconds after which S was presented with a letter of the alphabet. S was instructed to say the alphabet backwards from this start-point. After five seconds of alphabet-pronunciation, S was presented with a cue phrase which may or may not have appeared in one of the sentences in the preceding block of five. Half the time, the cue phrase was identical to a phrase presented in one of the preceding five sentences; half the time, the phrase had been altered in some manner.[5]

After the cue phrase was presented, S was asked to decide whether he had seen this phrase in the preceding five sentences (Yes/No).[6] S was also asked to enter a judgment of his confidence, on a scale from 1-5, concerning the recognition decision. If S felt he had seen the phrase, he was asked to attempt to recall as much of the critical sentence as possible.

Forty-eight blocks of sentences were presented to Ss. On half these trials, the cue phrase was identical to one of the phrases earlier presented; the phrase was "OLD." On the other 24 trials, the cue phrase was "NEW"--it represented a phrase that had been altered in some way.

Results and Discussion

The first question to be answered is the degree to which the confidence judgments given by the Ss reflect their true knowledge of the OLD/NEW status of the cue phrases. When OLD cue phrases are examined, it is found that an increase in the proportion of correct judgments (saying OLD when in fact the cue phrase is unchanged) increases smoothly from the confidence category of 1 up to the 5 category. The relevant proportions are presented in Table 4-1.

However, with NEW cue phrases, no such smooth relation obtains. Only with a confidence judgment of 5 is there a marked increase above an essentially flat linear trend for categories 1 through 4. It appears that the confidence judgments in the OLD condition more accurately reflect Ss' true knowledge of the OLD/NEW status of the cues (cf. Wearing, 1970).

Insert Table 4-1 about here

For this reason, it was decided not to include the confidence judgments in the analyses of the accuracy scores but rather to use a d' statistic (derived from signal detection theory) for the OLD cue phrases and a simple measure of proportion correct for the NEW condition.

OLD Cue Phrases

The d' measure of signal detectability may be considered as a type of correction for guessing or response bias. There are a number of possible acquiescence biases or response sets which may be operating in a recognition memory experiment of this sort. Included might be 1) an overall predisposition to answer OLD, 2) idiosyncratic response biases of Ss, and 3) bias towards OLD answers for particular sentence types. The d' statistic effectively counteracts these biases by treating the "OLD" answers (when the cue was in fact NEW) as a baseline noise level and then computing a measure of signal detectability for OLD cue phrases.

Each S proportions correct (when the cue phrase was OLD) and proportion incorrect (when the cue phrase was NEW) were converted to average d' scores. Each S obtained 3 d' scores, one for each of the three levels of the one repeated-measures factor in this analysis (Cue Phrase Presented). The other two factors included in the analysis were Concreteness and Voice.

There were two significant effects in this analysis. Concrete cue phrases were more easily detectable as OLD than abstract cue phrases ($F = 26.77$, $df = 1/28$, $p < .001$). As well, the two noun phrases (subject phrase and object

Table 4-1. Proportion of times correct at various confidence levels

	1	2	3	4	5
Old	.32	.53	.58	.62	.81
New	.72	.62	.70	.77	.85

phrase) were more often detected correctly as OLD than the verb phrase ($F = 5.99$, $df = 2/56$, $p < .005$). The failure to obtain a significant effect for Voice is surprising but is in line with Wearing (1970).

NEW Cue Phrases

The proportions correct for judging NEW phrases were analyzed in a 5-way analysis of variance where the factors were Voice, Concreteness, Cue Phrase Presented, Part of Speech Changed, and Type of Change.

Two significant main effects were obtained. Concrete phrases were correctly judged NEW more often (84%) than abstract phrases (69%; $F = 11.39$, $df = 1/28$, $p < .005$). Use of a synonym to replace a word in the cue phrase was a more effective distractor than use of an unrelated word (.67 vs .85 correct; $F = 34.72$, $df = 1/28$, $p < .001$). This effect has been obtained previously by Bruder and Silverman (1972), Anisfeld (1970), and Fillenbaum (1969).

There were three significant interactions. The means for the significant interaction of Type of Change by Cue Phrase Presented are presented in Table 4-2 and Fig. 4-1 ($F = 8.66$, $df = 2/56$, $p < .001$). This result seems to indicate that the subject phrase is relatively impervious to faulty recognition. However, both the verb phrase and the object phrase are prone to be perceived as OLD when a synonym is used as the distractor.

Insert Table 4-2 and Fig. 4-1 about here

The means for the significant ($F = 4.87$, $df = 2/56$, $p < .02$) interaction of Part of Speech Changed by Cue Phrase Presented are presented in Table 4-3 and Fig. 4-2.

Insert Table 4-3 and Fig. 4-2 about here

The significant interaction of Type of Change by Part of Speech Changed ($F = 4.67$, $df = 1/28$, $p < .05$) is presented in Table 4-4 and Fig. 4-3.

Insert Table 4-4 and Fig. 4-3 about here

Sentence Recall

For both OLD and NEW cue phrases, if S felt the cue was OLD, he almost invariably made some attempt to recall some of the associated sentence (98.8% of the time). However, there is some indication that even though on all sentence recall attempts S perceived the cue phrases as OLD, there was some differential recall. For truly OLD cue phrases, 8% of the sentence recall attempts produced only the cue phrase (or parts of it) while with NEW cue phrases (mistakenly

Table 4-2. Means for interaction of Cue Phrase Presented by Type of Change. Entries are mean proportions correct (NEW phrases).

		Type of Change	
		Synonym	Other
Cue Phrase Presented	Subject	.74	.79
	Verb	.56	.88
	Object	.72	.88

Fig. 4-1. Interaction of Cue Phrase Presented by Type of Change--
Mean proportions- correct recognition of NEW phrases.

Table 4-3. Mean proportions correct (NEW phrases) for interaction of Part of Speech Changed by Cue Phrase Presented.

		Part of Speech Changed	
		Modifier	Head
Cue Phrase Presented	Subject	.79	.74
	Verb	.74	.70
	Object	.74	.86

Fig. 4-2. Interaction of Cue Phrase Presented by Part of Speech Changed-- Mean proportions- correct recognition of NEW phrases.

Table 4-4. Mean proportions correct for interaction of Part of Speech Changed by Type of Change (NEW phrases)

		Part of Speech Changed	
		Modifier	Head
Type of Change	Synonym	.65	.70
	Other	.87	.83

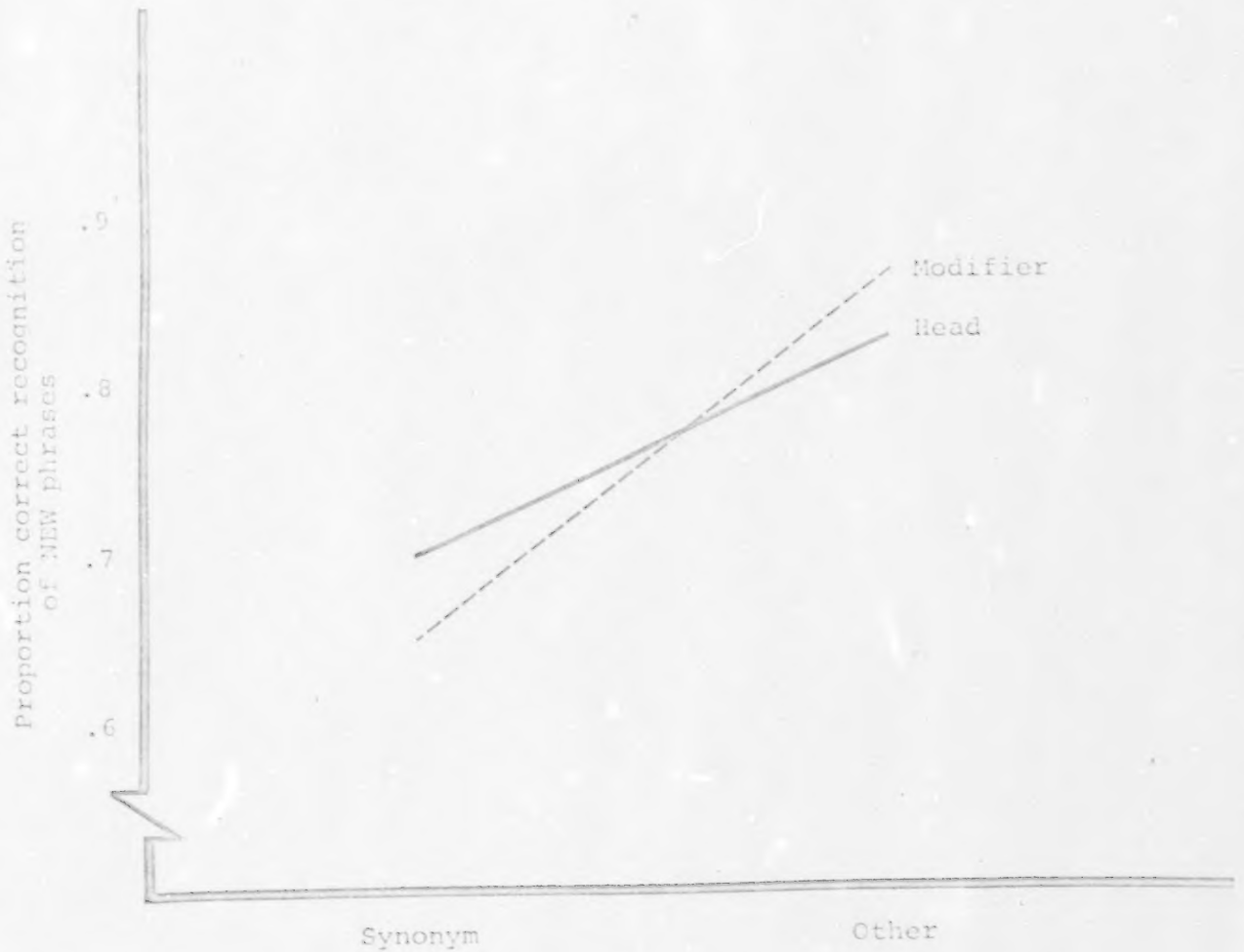

Fig. 4-3. Interaction of Part of Speech Changed by Type of Change-- Mean proportions- correct recognition of NEW phrases.

called OLD), 16.9% of the recall attempts were restricted to the presented cue phrase or parts of it.

However, almost always (98.5% of the time) whenever S made any recall attempt at all, S included some of the cue phrase in his response. In the analyses to follow the part of the recall attempt comprising the presented cue phrase will be deleted. As well, to avoid any artifactual results, the functor words that appeared constantly in every sentence ("the" and "was," and "by" in the passive form) were omitted when proportion of words recalled correctly was computed.

Only those instances where the cue phrase was unaltered (OLD), and S correctly perceived this, were included in the recall analysis. The design used for the analysis of variance was the same three-factor design used for the analysis of the signal detection scores for OLD cue phrases. There was a significant main effect for Cue Phrase Presented ($F = 5.23$, $df = 2/56$, $p < .009$). When the subject phrase was the cue, recall of the words of the rest of the sentence was 39% accurate, with the verb phrase as cue, 32% accuracy obtained; when the object phrase was cue, accuracy was only 27%. There was a significant main effect for Concreteness ($F = 15.35$, $df = 1/28$, $p < .001$). Concrete sentences were recalled better (42%) than abstract sentences (24%).

There was also a significant interaction between Voice and Concreteness ($F = 4.73$, $df = 1/28$, $p < .04$). This interaction will not be discussed in light of the significant three-way interaction between these two factors and Cue Phrase Presented ($F = 3.60$, $df = 2/56$, $p < .04$).

Insert Table 4-5 and Fig. 4-4 about here

Examination of Fig. 4-4 reveals the striking locus of the interaction effect. Three groups have essentially similar patterns; however, abstract passive sentences display a deviant pattern. For this type of sentence subject phrases function as superior cues. This suggests that the superiority for subject phrase cues noted above is due almost entirely to one type of sentence--abstract passive sentences. For concrete sentences and abstract active sentences each of the three phrases was reasonably equally efficacious as a cue. But for the abstract passive sentences the subject phrase clearly excels. Why is this? If, as has been hypothesized, concrete sentences are stored as global ideas, then no one cue phrase should be more important than any other for recall.

This does not seem to be the case for abstract sentences and particularly for abstract passive sentences. Here it appears that S focusses on the first phrase during presentation and treats the whole sentence in terms of this

Table 4-5. Interaction of Cue Phrase Presented by Concreteness by Voice--proportion of words recalled correctly.

		Voice			
		Active		Passive	
		Concrete	Abstract	Concrete	Abstract
Cue Phrase Presented	Subject	.43	.21	.34	.42
	Verb	.45	.13	.39	.20
	Object	.53	.20	.39	.26

Fig. 4-4. Interaction of Cue Phrase Presented by Concreteness by Voice-- Proportion of words recalled correctly.

phrase. This type of effect has been noted previously (Tannenbaum & Williams, 1968; Turner & Rommetveit, 1968). Although the relative use of active and passive sentences for emphasizing different types of information (e.g., actor or patient) is well-known, the present result is more complicated. The focus on the subject phrase which we might predict regardless of voice appears only for the passive voice for abstract sentences.

Misplaced Phrases

One of the questions which gave rise to this study involved the order of emission of the subject and object phrases in reversible sentences. Two predictions were made: 1) that since abstract sentences should be stored in a form more closely resembling the left-to-right order of surface structure, abstract sentence recall should manifest less confusability of placement of subject and object phrases at recall; and 2) there should be less confusability when the voice is passive and the cue phrase is the object phrase (due to the important element "by").

Neither of these predictins was supported by the results. Although there was a significant difference (F = 6.00, df = 1/28, $p < .02$) for probability of recall in the wrong phrase position between concrete and abstract sentences, concrete sentences showed less confusability of phrase placement (.05) than abstract sentences (.14).

The other main effect (Cue Phrase Presented; F = 8.39, df = 2/56, $p < .001$) showed that object phrase cues resulted in considerable confusion (.19) while subject and verb cues were less disruptive of order of recall (.04 and .06 for subject and verb cues, respectively).

This latter result is particularly interesting. Why does the object phrase result in so much disruption of order of recall of noun phrases when it is used as a cue? It is not simply because it is a noun phrase since the other noun phrase (subject phrase) did not have a similar effect. First, we know from the results above that when the object phrase is used as the cue, proportion of words correct is at its poorest. Thus, the object phrase is a poor probe for the entire sentence since it functions less effectively in eliciting the other components. This suggests that its relation to the other elements in memory storage is the weakest of the three phrases. If this is the case, then when the object phrase is presented as a cue, it acts as information which in some sense is less well integrated with the sentence than the other two components. This is the second part of the explanation. Since the object phrase is less attached, and it is the only information presented at recall test time, there is a greater likelihood that its position (as subject or object) will be more often confused.

At this point we should make some attempt to integrate our results in Exp. 4. Two related views of the dynamics of sentence learning assert that when prose is learned, the core meaning of the sentence is extracted and remembered (e.g., Gomulicki, 1956) or that surrogate structures are formed and stored. Pompi and Lachman (1967) proposed that these surrogate structures are based in part on the word order of the text and the associative relations among words but not necessarily containing any direct copies or representations of either the actual words or the grammatical structure of the sentence.

In Gomulicki's (1956) statement of the "abstraction of core meaning" type of theory, he posits that Ss omit unimportant words and remember an action-agent-effect unit. Gomulicki found that verbs were best remembered, then agents, then objects. Our results run counter to this finding. For OLD cues, NEW cues, and the recall date, the logical subject was always the most powerful part of the sentence. It was least subject to confusion in recall and functioned as the best retrieval cue. In all cases, except for the analysis of OLD cue phrases, object phrases were weakest. They were more often false identified when a slight change was made, they were the poorest retrieval cue, and when used as a cue, resulted in the greatest number of subject-object displacements at recall.

This latter result strongly questions the completeness of a model that predicts the recall of ordered information based only on the surface structure relationships existing in a sequence. There appears to be another, more cognitive mechanism operating. Given the pattern of results obtained, it seems that the sequential information contained in the sentences is stored in a surrogate form. Although the memory representation may be partly hierarchical (as suggested by Johnson), it seems more likely that it is the meaningful dynamic relationships among the idea constituents of the sentence that more influence recall.

Consistently throughout this study concrete sentences surpass abstract sentences on all measures of recallability. This finding is congruent with the earlier speculation--the critical process seems to be the method of memory storage and the relationships among the units in the store rather than the coding operations hypothesized to operate at retrieval.

General Conclusions

This series of experiments set out to explore the implications of Johnson's decoding-operation model of recall where the stimulus materials were meaningful English sentences of several types. Four different experimental paradigms were used: paired-associate learning, cued recall, recognition memory, and shared-capacity recall. In all these studies only slight correspondence to the results predicted by a theory that seeks to explain performance based only on the phrase-structure of the sentence was obtained. It may be useful to backtrack for a moment and re-analyze the assumptions under which these studies were planned.

It is possible to conceptualize performance in an experimental study of memory as involving at least three processes. First, the process of learning itself. Learning is probably not the proper term to use in the present context since we are assuming that all the lexical items present in the sentences are meaningfully available to S. Rather, S is asked to temporarily consider some of the items of his vocabulary more salient and make them more available for recall when asked to do so. As well, S is asked to retain certain order information concerning these lexical items.

The second process involves the storage of these two types of information. Here we had been working under the implicit assumption that the type of storage was at least partly constrained or determined by the phrase-structure chunk units defined by the grammatical form of the sentence.

Third, the process of recall or retrieval plays a very important role in performance. It is here that the decoding from memory storage occurs, and it is here that the decoding-operation model specifies a number of predictions. However, as was pointed out in Experiment 1, a number of these predictions fail to be confirmed when meaningful English sentences are used as the test materials. This is not to say that Ss do not make use of the phrase-structure of a sentence in comprehension and recall. This undoubtedly occurs (Johnson, 1965). The question we may ask at this point is: Is this all that occurs? A number of other processes seem to have more potency during the whole learning-recall process.

The results of Experiment 4 clearly demonstrate the probable role of storage factors. Although the phrase-structure cues may have been used by S during comprehension of the sentences, it is clear that the sentences were stored in some form other than as words.

Experiment 2 shows that there is little difference between age groups in how they organize outputs but clear

and striking differences that could be attributed to storage mode. Experiment 3 also obtained essentially negative results for manipulations designed to affect degree of chunking, with one exception. It appears that increasing the memory load to a certain point induces a chunking strategy on certain trials.

There is a possibility that a fourth set of processes may be operating in our experiments. The studies were not designed to test predictions related to this possibility so any theorizing of the role of reconstructive processes must remain merely speculation.

Reconstruction refers to the process by which S recreates the stimulus material at output in terms of his own conception of the world. This is most clearly seen in experiments involving complex cognitive material such as stories, but it may be expected to operate at the single-sentence level also. Reconstruction, of a kind, can also operate at the storage stage. Unfamiliar or unusual materials presented to Ss could be most efficiently stored as close approximations to already-known knowledge. Reconstruction can also be used as a partial explanation of the many effects described earlier for phrase-structure divisions. The output during a memory experiment may only coincidentally match the input because the rules used in the construction of the stimuli may be the same rules used by S in his construction of the response. S is not simply outputting the stored contents of memory--he is constructing an output (from a few stored pieces of information). The structure and form of the response appear to have been stored but in reality may not have been.

One clear set of findings in this series of studies need not rest on speculation. The variable of Concreteness-Abstractness has shown itself to be a very powerful one in all the experiments. Concrete sentences were consistently recalled better than abstract ones regardless of the mode of measurement of recall. In some cases the superiority of concrete materials is embarrassing--in Experiment 4. A general theory of storage that predicts that abstract sentences are stored with greater correspondence to word-to-word order fails to predict the more accurate recall of the order of large elements in concrete sentences even when a conspicuous attempt was made to confuse Ss by the use of reversible sentences. This finding cannot be interpreted by any current theory of memory. Even a reconstructive approach would fail to predict this effect since there is no a priori reason why concrete reversible sentences should be "closer to experience" in the presented form than similar abstract sentences.

In summary, the various results of the experiments lead us to a view of human recall (even in the relatively artificial situations used here) that must emphasize at

least four interacting processes--emphasis on phrase-structure and decoding operation alone provides only a partial picture of the processes at work. Doubtless, the phrase-structures of the presented stimuli act as ready-made chunks useful for comprehension and storage. It is clear, however, that semantic factors also influence storage. Finally, there is a hint that reconstructive processes may be used by S at recall time.

Footnotes

1. A different set of nine two-digit numbers was generated for each S.

2. The order of the two adjectives was counterbalanced across Ss.

3. Since there were two different adjectives associated with each sentence, when Ss were being tested in the one-adjective condition, for half the Ss one adjective was used, and for the other Ss the other adjective was presented.

4. Each target sentence was randomly selected from a pool of sentences; the filler sentences were selected from a second pool. Each possible filler sentence was examined to ensure that it had no words (other than "the") in common with the target sentence in that block.

5. The decision regarding whether a cue phrase should be the same or altered was performed randomly by the computer with the restriction that no one type be presented more than four times in succession.

6. While answering, S was prevented from seeing the block of five sentences by a cardboard screen attached to the top and rear of the teletype.

References

Anderson, B. F. The short-term retention of active and passive sentences. Unpublished doctoral dissertation, Johns Hopkins University, 1963.

Anglin, J., & Miller, G. A. The role of phrase structure in the recall of meaningful verbal material. Psychonomic Science, 1968, 10(10), 343-344.

Anisfeld, M. False recognition of adjective-noun phrases. Journal of Experimental Psychology, 1970, 86(1), 120-122.

Begg, I., & Paivio, A. U. Imagery and sentence meaning. Journal of Verbal Learning and Verbal Behavior, 1969, 8(6), 821-827.

Binet, A., & Henri, V. La memoire des mots. L'Annee Psychologique, 1894, 1, 1-59.

Boakes, R. A., & Lodwick, B. Short term retention of sentences. Quarterly Journal of Experimental Psychology, 1971, 23, 399-409.

Bruder, G., & Silverman, W. Effects of semantic and phonetic similarity on verbal recognition. Journal of Experimental Psychology, 1972, 94(3), 314-320.

Clark, H. H. Some structural properties of simple active and passive sentences. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 365-370.

Dolinsky, R., & Michael, R. Post-integration in the recall of grammatical and ungrammatical word sequences. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 26-29.

Epstein, W. The influence of syntactical structure on learning. American Journal of Psychology, 1961, 74, 80-85.

Epstein, W. A further study of the effect of syntactical structure on learning. American Journal of Psychology, 1962, 75, 121-126.

Epstein, W. Some conditions of the influence of syntactical structure on learning: grammatical transformation, learning instructions, and "chunking". Journal of Verbal Learning and Verbal Behavior, 1967, 6, 415-419.

Fillenbaum, S. Memory for gist: some relevant variables. Language and Speech, 1966, 9, 217-227.

Fillenbaum, S. Words as feature complexes: False recognition of antonyms and synonyms. Journal of Experimental Psychology, 1969, 82(2), 400-402.

Flagg, P., & Reynolds, A. G. A Z-statistic for sentence recall measurement. Department of Psychology, Dartmouth College, Mimeo, December, 1972.

Fodor, J. A., & Bever, T. G. The psychological reality of linguistic segments. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 414-420.

Furakawa, J., Suydam, M., & Miller, S. The effects of grouping instructions in short-term memory. Psychonomic Science, 1969, 15(2), 81-82.

Garrett, M., Bever, T., & Fodor, J. A. The active use of grammar in speech perception. Perception & Psychophysics, 1966, 1, 30-32.

Garrett, M., & Fodor, J. A. Psychological theories and linguistic constructs. In T. R. Dixon and D. L. Horton (Eds.), Verbal Behavior and General Behavior Theory. Englewood Cliffs, N. J.: Prentice-Hall, 1968, Pp. 450-477.

Gomulicki, B. R. Recall as an abstractive process. Acta Psychologica, 1956, 12, 77-94.

Halliday, M. Some aspects of the thematic organization of the English clause. Santa Monica: The Rand Corporation (Memorandum RM-5224-PR), 1967(a).

Halliday, M. Notes on transitivity and theme in English, Part II. J. Linguistics, 1967(b), 3, 199-244.

Hebb, D. O. On language and thought. Unpublished manuscript, Department of Psychology, McGill University, 1969.

Horowitz, L. M., & Prytulak, L. S. Redintegrative memory. Psychological Review, 1969, 76(6), 519-531.

Johnson, N. F. The psychological reality of phrase-structure rules. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 469-475.

Johnson, N. F. Sequential verbal behavior. In T. R. Dixon and D. L. Horton (Eds.), Verbal Behavior and General Behavior Theory. Englewood Cliffs, N. J.: Prentice-Hall, 1968, Pp. 421-450(a).

Johnson, N. F. The influence of grammatical units on learning. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 236-240(b).

Johnson, N. F. Chunking: Associative chaining versus coding. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 725-731.

Johnson, N. F. Chunking and organization in the process of recall. In G. H. Bower (Ed.), The Psychology of Learning and Motivation, Vol. 4. New York: Academic Press, 1970, pp. 171-247.

Johnson, N. F. Organization and the concept of a memory code. In A. Melton and E. Martin (Eds.), Coding Processes in Human Memory. Washington: Winston/Wiley, 1972.

Johnson-Laird, P. N. The interpretation of the passive voice. Quarterly Journal of Experimental Psychology, 1968, 20(1), 69-73.

Lachman, R., & Tuttle, A. Approximations to English and short-term memory: Construction or storage. Journal of Experimental Psychology, 1965, 70, 386-393.

Lambert, W. E., & Paivio, A. The influence of noun-adjective order on learning. Canadian Journal of Psychology, 1956, 10, 9-12.

Lockhart, R. S. Retrieval asymmetry in the recall of adjectives and nouns. Journal of Experimental Psychology, 1969, 79, 12-17.

Maclay, H., & Osgood, C. E. Hesitation phenomena in spontaneous English speech. Word, 1959, 15, 19-44.

Mandler, G. Organization and memory. In K. Spence and J. T. Spence (Eds.), The Psychology of Learning and Motivation, Vol. 1. New York: Academic Press, 1967, pp. 327-372.

Marks, L. Some structural and sequential factors in the processing of sentences. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 707-713.

Marks, L., & Miller, G. A. The role of semantic and syntactic constraints in the memorization of English sentences. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 1-5.

Martin, E., & Roberts, K. Grammatical factors in sentence retention. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 211-218.

Martin, E., & Roberts, K. H. Sentence length and sentence retention in the free learning situation. Psychonomic Science, 1967, 8, 535.

Martin, E., Roberts, K. H., & Collins, A. M. Short-term memory for sentences. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 560-566.

Matthews, E. W. Developmental changes in the functional knowledge of phrase structure rules. Unpublished Masters Thesis, The Ohio State University, 1965.

Matthews, W. A. Transformational complexity and short-term recall. Language and Speech, 1968, 11(2), 120-128.

McNulty, J. A. Measurement of "adopted chunks" in free recall learning. Psychonomic Science, 1966, 4, 71-72.

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 1963, 2(4), 346-351.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97(a).

Miller, G. A. Human memory and the storage of information. IRE Trans. Inform. Theory IT-2, 1956, 129-137(b).

Morton, J. Discussion of "On hearing sentences." In J. Lyons and R. J. Wales (Eds.), Psycholinguistics Papers. Edinburgh University Press, 1966, pp. 10-23.

Müller, G. E., & Pilzecker, A. Experimentelle Beitrage zur. Lehre vom Gedachtnis Zeitschrift fur Psychologie und Physiologie der Sinnesorgane, 1900, Ergbd. 1.

Paivio, A. Learning of adjective-noun paired associates as a function of adjective-noun word order and noun abstractness. Canadian Journal of Psychology, 1963, 17, 370-379.

Paivio, A. U. Imagery and Verbal Processes. New York: Holt, Rinehart, Winston, 1971.

Perfetti, C. A. Sentence retention and the Depth Hypothesis. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 101-104.

Philipchalk, R. P. Thematicity, abstractness, and the long-term recall of connected discourse. Psychonomic Science, 1972, 27(6), 361-362.

Prentice, J. L. Effects of cuing actor vs cuing object on work order in sentence production. Psychonomic Science, 1967, 8(4), 163-164.

Reynolds, A. G. Information processing when translating or transforming sentences. Unpublished doctoral dissertation, Department of Psychology, McGill University, 1970.

Roberts, K. Some effects of grammatical context on the recall of sentence elements. Paper presented at the Mid-western Psychological Association Convention, Chicago, 1967.

Roberts, K. H. Grammatical and associative constraints in sentence retention. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 1072-1076.

Rommetveit, R. Words, meanings, and messages. New York: Academic Press, 1968.

Rosenberg, S. Association and phrase structure in sentence recall. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 1077-1081.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2(9), 437-442(a).

Sachs, J. S. Recognition of semantic, syntactic, and lexical changes in sentences. Paper presented at Psychonomic Society, October, 1967(b).

Sasson, R. Y., & Fraisse, P. Images in memory for concrete and abstract sentences. Journal of Experimental Psychology, 1972, 94(2), 149-155.

Savin, H. B., & Perchonock, E. Grammatical structure and the immediate recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 348-353.

Simon, H. How big is a chunk? Paper presented at Eastern Psychological Association Meetings, 1970.

Tannenbaum, P. H., & Williams, F. Generation of active and passive sentences as a function of subject or object focus. Journal of Verbal Learning and Verbal Behavior, 1968, 7(1), 246-250.

Thompson, C. P., & Roenker, D. L. Learning to cluster. Journal of Experimental Psychology, 1971, 91(1), 136-139.

Thorndike, E. L. Human Learning. New York: Century, 1931.

Thorndike, E. L., & Lorge, I. The Teacher's Word Book of 30,000 Words. New York: Teachers College Press, 1944.

Tulving, E., & Patkau, J. E. Concurrent effects of contextual constraint and word frequency on immediate recall

and learning of verbal material. Canadian Journal of Psychology, 1962, 16, 83-95.

Turner, E. A., & Rommetveit, R. Focus of attention in recall of active and passive sentences. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 543-548.

Uhlenbeck, E. M. An appraisal of transformation theory. Lingua, 1963, 12, 1-18.

Wanner, H. E. On remembering, forgetting, and understanding sentences: A study of the deep structure hypothesis. Unpublished doctoral dissertation, Harvard University, 1968.

Wearing, A. J. The storage of complex sentences. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 21-29.

Wood, G. Retrieval cues and the accessibility of higher-order memory units in multitrial free recall. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 782-786.

Wright, P. Sentence retention and transformation theory. Quarterly Journal of Experimental Psychology, 1968, 20(3), 265-272.

Yngve, V. H. A model and an hypothesis for language structure. Proceedings of the American Philosophical Society, 1960, 104, 444-466.